

"Phycocyanin exprimierende Eukaryontenzelle"

Die Erfindung betrifft rekombinant hergestelltes Phycocyanin bzw. funktionell aktive Teile davon und nimmt die Priorität der europäischen Patentanmeldung 04 001 504.2 in Anspruch, auf die inhaltlich Bezug genommen wird..

5 In allen Bereichen der modernen Biologie - z.B. bei der DNA-Replikation, Transkription, Translation oder Kontrolle des Zellzyklus sowie Signaltransduktionen spielt die Wechselwirkung zwischen Proteinen eine essentielle Rolle. Konsequenterweise ist die Aufklärung dieser Abläufe nicht mehr nur allein die Domäne von Biochemikern. Vielmehr müssen sich nun auch Genetiker sowie Molekular-  
10 biologen oder Biophysiker mit der Aufklärung von Protein-Protein-Wechselwirkungen beschäftigen.

Die Interaktionen zwischen Proteinen werden klassischerweise mit Hilfe biochemischer Methoden wie der Affinitätschromatographie oder der Immunpräzipitation  
15 identifiziert oder quantifiziert. Häufig werden diese Methoden mit genetischen bzw. molekularbiologischen Verfahren kombiniert. Wird beispielsweise ein gesuchtes Protein von einer Bakteriophagen/cDNA Bibliothek exprimiert, kann die Interaktion in manchen Fällen mit einer spezifischen Protein-Sonde direkt im Lysat eines positiven Plaques nachgewiesen werden.

20 Eine Vielzahl der Verfahren zur Untersuchung von Protein-Protein-Wechselwirkungen beruht jedoch auf der Detektion dieser Wechselwirkungen *in vitro*. So ist zum einen die Übertragbarkeit der damit erzeugten Ergebnisse auf *in vivo* Bedingungen fraglich. Des weiteren können nur Wechselwirkungen detektiert werden  
25 von Bindungspartnern, deren Dissoziationskonstanten nicht allzu groß sind. Anderenfalls würden die zu detektierenden Wechselwirkungen den in der Regel stringenten und zeitaufwendigen Waschprotokollen nicht standhalten. Diese sind aber üblicherweise notwendig, um die störenden Hintergrundsignale auf ein akzeptables Maß senken zu können. Die Entwicklung von Systemen zur Untersuchung der  
30 Protein-Protein-Wechselwirkungen *in vivo* stellt daher eine besondere Herausfor-

CONFIRMATION COPY

derung dar.

In der jüngeren Vergangenheit verwenden viele Wissenschaftler zur Untersuchung der Zellbiologie das sogenannte grüne Fluoreszenzprotein (GFP) aus *Aequorea victoria*. Bei GFP handelt es sich um ein fluoreszierendes Protein, dessen Klonierung und Expression in der Zwischenzeit in einer Vielzahl von Zellen und Organismen – z. B. in Bakterien, Pflanzen- und Tierzellen – gelungen ist. Damit kann das GFP in der Zellbiologie beispielsweise für die Lokalisation von Proteinen, für den Nachweis einer erfolgreichen Transfektion oder als Reportergen für eine Genexpression dienen.

Das GFP zeigt ein Absorptionsspektrum mit zwei Peaks, nämlich einen ersten Peak bei 395 nm und einen zweiten bei 470 nm. Dafür ist das Tyr<sup>66</sup> verantwortlich, das in seiner Hydroxy-Form durch die Anregung bei 395 nm in die Phenolat-Form übergeht. Diese emittiert nach entsprechender Anregung Licht einer Wellenlänge von 509 nm.

Damit liegt das Emissions- und Absorptionsspektrum des GFP genau in den spektralen Bereichen, in denen eine Vielzahl biologischer Substanzen eine Fluoreszenz aufweisen. Diese natürliche Fluoreszenz biologischer Systeme – auch als Autofluoreszenz bekannt – geht zu nicht unerheblichen Teilen auf Flavinderivate oder auf NADH- sowie NADPH-Derivate zurück. So werden Flavine beispielsweise durch Licht von 380-490 nm angeregt und emittieren Licht einer Wellenlänge von 520-560 nm. NADH sowie NADPH zeigen Absorptionsmaxima bei 360-390 nm, während sie die höchste Emission bei 440-470 nm aufweisen.

Diese Autofluoreszenz überschneidet sich daher mit dem für das GFP Protein typischen Spektrum, und ist bei der Verwendung von GFP in *in vivo* Studien für ein hohes Hintergrundrauschen verantwortlich. Dieses Hintergrundrauschen führt nicht selten zu einem Verlust an Klarheit und Aussagekraft der Versuchsergebnisse.

In der Vergangenheit wurden Verfahren entwickelt, um die Autofluoreszenz von der GFP Fluoreszenz zu unterscheiden. Diese Verfahren beruhen beispielsweise auf der Optimierung der für die Fluoreszenzdetektion verwendeten optischen Filter oder auf der sequentiellen Anregung der Probe zunächst mit einer auf die Anregung des GFP optimierten Wellenlänge und im Anschluß daran mit einer auf die Autofluoreszenz optimierten Wellenlänge. Die Differenz zwischen den erhaltenen

Meßsignalen gilt als Fluoreszenz des GFP allein. Diese letzte Methode ist als "Dual Wavelength Correction" bekannt. Auch andere Verfahren, so beispielsweise die "Microscopy Image Correction" basiert letztlich auf einer Korrektur des GFP Fluoreszenzsignals um den Wert der Autofluoreszenz.

5

Daneben wurden aufwendige Optimierungen an der eigentlichen Auswerteeinheit vorgenommen. So wird beispielsweise die konfokale Lasermikroskopie oder die Zwei-Photonen-Laserscan-Mikroskopie eingesetzt, um eine hohe Auflösung der Fluoreszenzsignale zu erhalten und sie somit möglichst einzelnen konkreten drei-  
10 dimensionalen oder räumlichen Strukturen innerhalb der biologischen Probe zuzuordnen zu können. Dieser letzte Ansatz geht auf Beobachtungen zurück, dass die Autofluoreszenz innerhalb der Probe meist nicht homogen verteilt ist.

Ein Großteil der Überlegungen zur Verbesserung der Auswertung eines GFP-Signals in einer *in vivo* Probe setzt daher an der Gestaltung des Messverfahrens sowie an einer Optimierung der Detektions- oder Auswerteeinheit an. (siehe dazu ausführlich: Billinton N und A.W. Knight: Seeing the Wood through the Trees: A Review of Techniques for distinguishing GFP from Cellular endogenous Autofluorescence. Analytical Biochemistry 2001 (291), 175-197).

20

Daneben existieren nur wenige Ansätze, das GFP durch Mutagenese so zu verändern, dass die Absorption und die Emission ins längerwellige Licht verschoben wird und somit den kritischen Bereich der Autofluoreszenz verlässt. So wurde z. B. das sogenannten Red Fluorescent Protein (RFP oder DsRed) aus *Discosoma* mit  
25 einem Emissionsmaximum bei 583 nm isoliert, dessen Weiterentwicklung mRFP1 ein Absorptionsmaximum bei 619 nm zeigt. Diese Variante hat zwar - im Gegensatz zum ursprünglichen RFP - keinen störenden zweiten Peak bei 490 nm. Es ist aber durch eine demgegenüber geringere Photostabilität und Signalstärke gekennzeichnet (Zhang J., E. Campbell, A. Ting und Y. Tsien: Creating New Fluorescent Probes for Cell Biology. Molecular Cell Biology 2002 (3), S. 906-918).

30

Aufgabe der vorliegenden Erfindung ist es daher, ein alternatives Fluoreszenzprotein insbesondere für zellbiologische Untersuchungen, sowie ein Verfahren zu seiner Herstellung bereitzustellen. Dieses Protein sollte insbesondere ein Absorptionsmaximum im längerwelligen Bereich aufweisen.

35

Gelöst wird diese Aufgabe durch die Bereitstellung eines Phycocyanins oder funktionell aktiver Teile davon, wobei das Phycocyanin bzw. dessen erfindungsge-

mäße Teile im Vergleich zum nativen Phycocyanin eine Glycosylierung aufweisen. Des weiteren wird eine transformierte Eukaryontenzelle, insbesondere eine Hefezelle, bereitgestellt, die das erfindungsgemäße Phycocyanin bzw. dessen funktionell aktive Teile exprimiert.

5

Bei Phycocyaninen handelt es sich um Phycobiliproteine, die als akzessorische Photosynthesepigmente aus Cyanobakterien, Rotalgen oder Cryptomonaden bekannt sind. In Cyanobakterien und Rotalgen liegen einzelne Phycobiliproteine überwiegend aggregiert zu großen Untereinheiten, den sogenannten Phycobilisomen, vor.

10

Die Phycobiliproteine gehören zu der Familie homologer Proteine, die jeweils aus einem eng miteinander verbundenen  $\alpha\beta$ -Heterodimer bestehen. Beide Untereinheiten - also sowohl die  $\alpha$ - als auch die  $\beta$ -Untereinheit können ein oder mehrere Chromophoren enthalten. Die Qualität und Anzahl der kovalent an die Proteineinheit (Apo-Protein) gebunden Chromophore dient der Klassifizierung der Phycobiliproteine.

15

Die Phycocyanine tragen eine einzige Bilin-Einheit an der  $\alpha$ -Untereinheit und zwei Biline an der  $\beta$ -Untereinheit. Im Vergleich dazu sind beispielsweise Phycoerythrine durch zwei oder drei Biline an der  $\alpha$ -Untereinheit und drei an der  $\beta$ -Untereinheit charakterisiert. Allophycocyanine hingegen tragen an jeder Proteinuntereinheit nur jeweils ein Bilin.

20

Wie bereits erwähnt, werden die Biline aufgrund ihrer Struktur unterschieden. Bei allen Bilinen handelt es sich jedoch übereinstimmend um Chromophore aus einem linearen Tetrapyrrol, das über eine Thioetherbindung zwischen dem C3'-Atom des Bilins und einem Cysteinrest der Proteineinheit kovalent an das Apo-Protein gebunden ist. Die Anzahl der Thioetherbindungen kann jedoch variieren. So weisen einige Biline zwei Thioetherbindungen zum Apo-Protein auf. Ebenso kann die Konfiguration des C3'-Atoms R oder S sein.

25

30

Die spektroskopischen Eigenschaften der Phycobiliproteine werden wesentlich durch die Anzahl und die Lage der Doppelbindungen in dem Bilin bestimmt. So wird zwischen dem Phycocyanobilin (PCB), dem Phycobiliviolin (PXB), dem Phycoerythrobilin (PEB) und dem Phycourobilin (PUB) unterschieden. Die zwei Thioetherbindungen zum Apo-Protein aufweisenden PEB und PUB tragen die Bezeichnung DiCysPEB bzw. DiCysPUB.

35

Daher ist die Konformation des Bilin-Anteils für die spektralen Eigenschaften des Phycobiliproteins von großer Bedeutung. So zeigen Röntgenstrukturanalysen, dass das Bilin im Phycobiliprotein in der Regel eine möglichst gestreckte Konformation einnimmt. Liegt das Bilin dagegen frei in Lösung vor – wodurch es eine davon abweichende Konformation einnehmen kann - verschiebt sich das Absorptionsmaximum stärker in den UV-Bereich.

Das Phycocyanin ist – wie bereits erwähnt – durch ein Bilin an der  $\alpha$ -Untereinheit und zwei Biline an der  $\beta$ -Untereinheit gekennzeichnet. Bei dem Bilin der  $\alpha$ -Untereinheit handelt es sich stets um ein PCB. Die  $\beta$ -Untereinheit kann entweder zwei PCB oder aber ein PCB und PEB tragen. Im ersten Fall handelt es sich um das C-Phycocyanin (C-PC), während das zweite Molekül als R-Phycocyanin (R-PC) bezeichnet wird. Die beiden PC-Varianten unterscheiden sich allerdings in ihren spektralen Eigenschaften nicht wesentlich.

Die meisten Phycobiliproteine absorbieren im Bereich von 495-650 nm. Phycocyanin zeigt – je nach Literatur, Versuchsbedingungen und Ausgangsorganismus - ein Absorptionsmaximum bei 600-650 nm. Nach Tooley *et al* (Tooley A.J., Yuping A.C., Glazer A.N.: Biosynthesis of a fluorescent cyanobacterial C-Phycocyanin holo- $\alpha$  subunit in a heterologous host. PNAS 2001 Vol 19) 10560 – 10565) zeigt das Holo-CpcA (d.h. also die  $\alpha$ -Untereinheit des PC) ein Absorptionsmaximum von 619 nm und eine maximale Emission bei 641 nm.

Diese spektralen Eigenschaften des Phycocyanins erlauben eine klare Diskriminierung des eigentlichen Fluoreszenz-Meßsignals von dem störenden Hintergrundrauschen der Autofluoreszenz und begründen somit dessen besondere Eignung als Fluoreszenzmarker für *in vivo*-Untersuchungen. Die Autofluoreszenz beruht - wie bereits oben erwähnt - im wesentlichen auf Flavin- oder NADH- bzw. NADPH-Derivaten, die Absorption- und Emissionsmaxima bei 360-560 nm zeigen.

Die grundlegenden Untersuchungen zur Biosynthese des Phycocyanins wurden an der Rotalge *Cyanidium caldarium* gemacht. Dabei wurde durch  $^{14}\text{C}$ -Markierungen des Häms deutlich, dass die Biosynthese des Phycocyanins zunächst die Degradation des endogenen Häms zu Biliverdin IX $\alpha$  durch die Hämoxygenase 1 (HO1) voraussetzt. Dabei wird der Tetrapyrrolring linearisiert. In weiteren Schritten wird das Biliverdin zu 14,16 Dihydrobiliverdin IX $\alpha$  und anschließend zu 3 Z - Phycocyanobilin umgewandelt.

Diese letzten beiden Umwandlungen setzen jeweils die Gegenwart von reduziertem Ferredoxin voraus und werden durch die Ferredoxin Oxidoreduktase katalysiert. Das so entstandene PCB wird durch die heterodimere Phycocyanin  $\alpha$  PCB Lyase an die  $\alpha$ -Untereinheit des Phycobiliproteins (Apo-CpcA) geknüpft. Damit entsteht die fertige Holo- $\alpha$ -Untereinheit (Holo-CpcA)

Die Biosynthese des Holo-CpcA umfaßt somit zunächst die Degradation und Umwandlung des Häms zum PCB und anschließend dessen enzymatische Bindung an die Proteinuntereinheit. Entsprechend verläuft die Biosynthese der Holo- $\beta$ -Untereinheit (Holo-CpcB), wobei die Bindung der Bilin-Einheiten an das CpcB vermutlich zumindest teilweise nicht-enzymatisch erfolgt. Liegen sowohl das Holo-CpcA als auch das Holo-CpcB vor, dimerisieren diese zu einem "Monomer" des Phycocyanins.

Die vollständige Biosynthese des Holo-CpcA wurde erstmals in der photosynthetisierenden Blaualge *Synechocystis sp.* PCC 6803 erkannt und die dafür relevanten Gene identifiziert (Nakamura, Y., Kaneko T., Hirose M., Miyajima N., Tabata S. (1998) Nucleic Acids Res. 26, 63-67). Die heterodimere PCB Lyase wird von den Genen *cpcE* und *cpcF* kodiert. Die Hämoxygenase wird durch *hox1* und die Ferredoxin Oxidoreduktase durch das *pcyA* Gen kodiert. Das *cpcA*-Gen kodiert die  $\alpha$ -Untereinheit CpcA und das *cpcB*-Gen die  $\beta$ -Untereinheit CpcB des Phycocyanins.

Trotz des großen Bedarfs an Fluoreszenzmarkern mit den besonderen spektralen Eigenschaften des Phycocyanins für die Zellbiologie und der Kenntnis der an der Biosynthese beteiligten Proteine sowie deren Gene ist es bisher nicht gelungen, Phycocyanin oder funktionell aktive Teile daraus – wie beispielsweise das Holo-CpcA oder das Holo-CpcB - rekombinant in Eukaryonten zu exprimieren.

In der Vergangenheit ist zwar die heterologe Expression des Holo-CpcA in *E. coli* gelungen (Tooley et. al. a.a.O.). Dazu wurden die Zellen mit zwei Plasmiden mit *cpcA*, *cpcE* und *cpcF* Gen bzw. *hox1* und *pcyA* transfiziert. Nach Induktion der Genexpression erfolgte die Biosynthese des Holo-CpcA. Auch zeigte dieses Holo-CpcA vergleichbare spektrale Eigenschaften wie das native Holo-CpcA. Die Ausbeute an Holo-CpcA war aber gering, da nur etwa ein Drittel des verfügbaren Apo-CpcA tatsächlich zum Holo-CpcA umgesetzt wurde.

In ihren Untersuchungen haben Tooley *et al.* evaluiert, ob möglicherweise ein Mangel an verfügbarem endogenen Häm in *E. coli* für die reduzierte Konversion von Apo-CpcA zu Holo-CpcA ursächlich ist. Dazu haben sie den transformierten Zellen zusätzlich  $\delta$ -Aminolävulinsäure angeboten. Die zusätzliche Gabe dieser  
5 Schlüsselverbindung des Häm-Stoffwechsels hat jedoch nicht zu einer Erhöhung der Holo-CpcA Ausbeute geführt.

Die Autoren spekulieren daher darüber, ob die für die Expression in *E. coli* ungünstige Codonverwertung ("Codon Usage") der Gene *hox1*, *pcyA*, *cpcE* und *cpcF*  
10 für die mangelhafte Ausbeute verantwortlich sein könnte. Des weiteren könnte eine ungleiche Anzahl der jeweils in einer *E. coli*-Zelle vorliegenden Plasmide - die entweder *hox1* und *pcyA* oder aber *cpcA*, *cpcE* und *cpcF* tragen - eine Rolle spielen (siehe dazu auch Tooley *et al.* a. a.O. Seite 10564 rechte Spalte oben).

15 Auch andere Arbeitsgruppen haben bereits vor langer Zeit die Vorteile der Verwendung von Phycobiliproteinen als Marker für intrazelluläre Vorgänge *in vivo* erkannt und daher versucht, Systeme zur Bereitstellung von Holo-CpcA in Hefen zu etablieren. Dazu wurden beispielsweise die Gene *cpcE* und *cpcF* für die heterodimere PCB-Lyase sowie das *cpcA*-Gen zu Kodierung des Apo- $\alpha$ -CPCA in die Hefen kloniert. Durch eine anschließende Gabe exogenen PCB war beabsichtigt, ein  
20 funktionelles Holo CpcA - das heißt also einen funktionell aktiven Teil des Phycocyanins - in den Hefen zu generieren. Dieser Versuch der semisynthetischen Bereitstellung des Holo-CpcA in Hefen schlug jedoch fehl. (Schröder B.G. (1997) Doctorial Dissertation (University of California, Berkeley): Phycobiliprotein: Biosynthesis and Applications)  
25

Die möglichen Ursachen für den fehlgeschlagenen Versuch, funktionell aktive Teile des Phycocyanins durch die Addition exogenen PCBs an rekombinant erzeugtes Apo-CpcA in Hefen zu generieren, wird von Schröder ausführlich diskutiert. So konnte er durch die Wahl geeigneter Versuchsbedingungen ausschließen,  
30 dass die fehlende Addition des PCB an CpcA durch einen möglicherweise fehlenden Import des externen PCB in die Zelle oder aber durch einen aktiven Export des zunächst importierten PCB aus der Zelle heraus bedingt war.

35 Er schloß aufgrund seiner Untersuchungen daher aus, dass die Additionsreaktion der in das Hefegenom einklonierten *CpcE* *CpcF* Lyase aufgrund einer unzureichenden Konzentration an PCB ausblieb. Er folgerte aus seinen Untersuchungen des weiteren, dass in der Hefe ein die Additionsreaktion hemmender Inhibitor vor-

handen sein muß, und verweist für die Zukunft auf die Möglichkeit, Hefemutaten zu identifizieren, die diesen Inhibitor nicht aufweisen.

Ferner spekuliert Schröder auf die Möglichkeit, nach einer Aufklärung des vollständigen Biosynthesewegs die dafür relevanten Gene in die Hefe einzuklonieren und somit eine vollständige *in vivo* Biosynthese zu erreichen (siehe zu der Diskussion dieses Problems auch Schröder a. a. O. Seite 77-79). Eine eindeutige Erklärung für das Scheitern seiner Bestrebungen zur Bereitstellung einer semi-synthetischen Biosynthese bleibt Schröder aber schuldig.

In diesem Zusammenhang sei nochmals darauf hingewiesen, dass es Tooley et al. aber selbst nach der Aufklärung des Biosynthesewegs nicht gelungen ist, eine zufriedenstellende Expression in *E. coli* zu erreichen (siehe oben).

Vor dem Hintergrund dieses Stands der Technik ist es um so überraschender, dass nun erstmals die Transformation einer Eukaryontenzelle gelungen ist, die ein Holo-CpcA als funktionell aktives Teil eines Phycocyanins exprimiert. Dabei wird in diesem Zusammenhang unter einem "funktionell aktiven Teil" jedes Teil, Analogon oder Derivat des Phycocyanins verstanden, das zu mindestens einer der Phycocyanin  $\alpha$ - oder  $\beta$ -Untereinheiten (oder Teilen davon) eine strukturelle Homologie aufweist und in seinen spektralen Eigenschaften im wesentlichen denjenigen Eigenschaften des nativen Phycocyanins entspricht.

In einer besonders bevorzugten Ausführungsform exprimiert die erfindungsgemäße Eukaryontenzelle das Holo-CpcA. Dazu kann die Eukaryontenzelle vorzugsweise mit den Genen *cpcE* und *cpcF* für die PCB-Lyase, dem *hox1* für die Hämoxygenase, dem *pcyA* für die Ferredoxin-Reduktase und dem *cpcA* für die  $\alpha$ -Untereinheit transformiert sein. Damit kann eine vollständig endogene Biosynthese des Holo-CpcA in einem heterologen eukaryontischen Wirt erreicht werden.

Ebenso ist es aber auch möglich, das Holo-CpcA über einen semisynthetischen Biosyntheseweg bereitzustellen. Dazu kann es ausreichen, das CpcA oder CpcB – oder zumindest den jeweils für die Addition des PCB relevanten Teils davon – sowie die PCB Lyase in der Eukaryontenzelle zu exprimieren und PCB exogen hinzuzuführen. Alternativ können auch ein oder mehrere der anderen für die Biosynthese relevanten Gene entweder einkloniert werden oder aber ihr jeweiliges Genprodukt extern hinzugefügt werden. Damit ist es möglich – in Abhängigkeit von der



jeweiligen konkreten Fragestellung des Versuchs — das System gezielt zu induzieren. Mit der anschließenden Isolierung des Holo-CpcA oder Holo-CpcB aus der Wirtszelle steht das erfindungsgemäße "funktionelle aktive Teil" des Phycocyanins bereit.

In dem Fall, dass sowohl das Holo-CpcA als auch das Holo-CpcB in der Zelle exprimiert werden — entweder vollständig endogen oder aber semisynthetisch wie oben beschrieben — können die beiden Untereinheiten dimerisieren und ein komplettes Phycocyanin-Monomer bilden. Erfindungsgemäß kann somit nicht nur ein Teil des Phycocyanins, sondern auch ein vollständiges Phycocyanin bereitgestellt werden.

Das erfindungsgemäß bereitgestellte Phycocyanin bzw. dessen funktionell aktive Teile weisen ein Glycosylierungsmuster auf, das sie von dem jeweils nativen Protein oder der entsprechenden Proteinuntereinheit unterscheidet. So ist beispielsweise das in einem eukaryontischen Wirt exprimierte CpcA an dem Threonin der Position 6, dem Serin an Position 10 und dem Serin an Position 162 der Aminosäuresequenz über eine O-Verknüpfung glykosyliert. Diese Glykosylierungen liegen im nativen CpcA nicht vor.

Ein besonderer Vorteil des erfindungsgemäßen Fluoreszenzproteins besteht darin, dass sein Emissionsspektrum im Vergleich mit seinen bisher bekannten Expressionsformen eine Rotverschiebung zeigt. Bei einem pH von 8,0 liegt das Emissionsmaximum des Holo-CpcA bei 643 nm und damit um 2 nm weiter im längerwelligen Bereich als beispielsweise das in *E.coli* exprimierte CpcA. Selbst wenn dieser Unterschied quantitativ nicht groß ist, kann er in der Praxis aber dazu dienen, dieses Signal von Signalen anderer Fluoreszenzfarbstoffe zu unterscheiden.

In einer besonders vorteilhaften Ausführungsform der Erfindung wird das Holo-CpcA in einer transformierten Hefezelle exprimiert. Dazu eignet sich insbesondere der unter der Aufnahmenummer DSM 16134 am 15. Januar 2004 bei der Deutschen Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) hinterlegte Hefestamm.

Es hat sich gezeigt, dass die heterologe Expression von Holo-CpcA nicht nur abhängt von der erfolgreichen Klonierung der für die Biosynthese erforderlichen Gene in die Wirtszelle, sondern auch durch eine geeignete Wahl der Kultivierungsbedingungen der Wirtszelle beeinflussbar ist. Dabei können insbesondere

Kultivierungsbedingungen eine Rolle spielen, die die Konzentration an verfügbarem endogenen Häm in der Wirtszellen beeinflussen. So kann eine Stimulierung der Häm-Biosynthese oder aber eine Hemmung des zellulären Häm-Abbaus zu einer Erhöhung der endogenen Häm-Konzentration führen.

5

Die insgesamt 8 Schritte umfassende Biosynthese des Häms ist seit langem bekannt und die dafür in der Hefe verantwortlichen Gene sind bereits identifiziert (z.B. R. Labbe-Bois, P. Labbe: Tetrapyrrol and Heme biosynthesis in the yeast *Saccharomyces cerevisiae*. In: H.A. Dailey (Ed.) Biosynthesis of Heme and Chlorophylls, Mc Graw-Hill, New York, 1990, 235-286 oder A. Chelstowska, J. Rytko: Heme biosynthesis in the yeast *Saccharomyces cerevisiae*. Postpy Biochem. 39 (1993) 173-185). Die Regulationsmechanismen dieses Syntheseweges sind jedoch noch nicht vollständig verstanden. Es hat sich aber gezeigt, dass die Produktionsrate des Häms u.a. von der verwendeten Kohlenstoffquelle im Medium und der Sauerstoffverfügbarkeit während der Kultivierung abhängt. So führt die Gabe von Glucose zu einer Repression der Hämsynthese. Gleiches gilt für andere von der Hefe fermentierbare Kohlenstoffquellen wie z.B. Galactose. Die Gabe von nicht-fermentierbaren Kohlenstoffquellen wie Ethanol stimuliert dagegen im Vergleich zu Glucose den Hämstoffwechsel um das 2 - 3-fache (M. Hoffmann, M. Gora, Rythka J: Identification of rate-limiting steps in yeast heme biosynthesis. Biochemical and Biophysical research Communications 310 (2003) 1247-1253; Chelstowska (1993)).

Somit ist es für die erfindungsgemäße Bereitstellung des rekombinanten Phycocyanins bzw. dessen Teile vorteilhaft, die transformierten Hefezellen in einem Medium mit nicht-fermentierbaren Kohlenstoffquellen zu kultivieren, um die endogene Konzentration des verfügbaren Häms zu steigern.

Alternativ können in einer vorteilhaften Ausführungsform des erfindungsgemäßen Herstellungsverfahrens Inhibitoren des Hämabbaus dem Kulturmedium zugesetzt oder in den Wirtsorganismus kloniert werden.

Bei dem Enzym HMX1 handelt es sich um ein Homolog der Hämoxygenase, das der unveränderten Hefe fehlt (Protchenko O., Philpott C.C. (2003) Regulation of Intracellular Heme Levels by HMX1, a Homologue of heme Oxygenase, in *Saccharomyces cerevisiae*. JBC 278:36582-36587). Da dieses Enzym über Eisen gehemmt wird, besteht die Möglichkeit, dieses Enzym in einem konkreten Kulturmedium zu hemmen, z. B. durch die Gabe von Eisen.

35

Die Konzentration an verfügbarem Häm kann erfindungsgemäß ebenso durch eine Überexpression der Schlüsselenzyme der Hämsynthese erfolgen. Da zu kann die Porphobilinogen Synthase, die Porphobilinogen Deaminase oder die Uroporphyrinogen III Decarboxylase überexprimiert werden (Hoffmann et al (2003) a.a.O.).

Ebenso kann es – bei der Verwendung des ADH-Promotors als Promotor der einklonierten Gene – vorteilhaft sein, die Aktivität des Promotors z.B. durch eine Erhöhung der Glucose-Konzentration zu steigern. Die Erhöhung der Glucosekonzentration zur Steigerung der Promotoraktivität steht zwar zunächst im Konflikt mit der damit verbundenen Repression des Hämstoffwechsels. Es hat sich aber experimentell bestätigt, dass insbesondere eine Erhöhung der Glucosekonzentration im Medium von dem bei der Kultivierung von Hefen üblichen 2 % auf beispielsweise 13 % zu einer optimierten Ausbeute von rekombinantem Holo-CpcA mit einer Steigerung des Fluoreszenzsignals um das mehr als 7-fache führt (siehe unten).

In einer besonders vorteilhaften Ausführungsform wird das Kulturmedium mit 2 % Ethanol versehen. Es ist bekannt, dass Ethanol einerseits als nicht fermentierbare Kohlenstoffquelle den Hämstoffwechsel induziert (Hoffmann et al (293) a. a. O.), aber andererseits auch die Aktivität des ADH Promotors fördert (Ruohon en, L., Aalto, M.K., Keränen, S.: Modifications to the ADH1 promotor of *Saccharomyces cerevisiae* for efficient production of heterologous proteins. *Journal of Biotechnology* 1995 (39) 193-203)a.a.O.).

Erfindungsgemäß kann alternativ oder ergänzend vorgesehen werden, den Abbau des rekombinanten Phycocyanins in der Zelle bzw. dessen Teile zu hemmen oder zu verzögern. Da ein Stickstoffmangel einen verstärkten Abbau der Phycocyanin Hexamere begünstigt (Grossman AR (1990) Chromatic adaptation and the events involved in phycobilisome biosynthesis. *Plant, Cell and Environment* 13:651 -666), kann eine Supplementierung mit Stickstoff dem Phycocyanin-Abbau entgegenwirken.

Wie bereits erwähnt, wird in einer Ausführungsform der Erfindung eine Hefezelle transformiert. Dazu ist es vorteilhaft, die einklonierten PC-Gene für die Expression in der Hefe zu optimieren, indem zunächst der GC-Gehalt der Gene auf einen durchschnittlichen Wert von 45% gebracht wird. So kann die genetische Stabilität

und die Halbwertszeit der mRNA im Wirtsorganismus verlängert werden.

Darüber hinaus kann die Effizienz der Expression noch gesteigert werden, indem die PC-Gene hinsichtlich ihrer Codon-Verwertung ("Codon Usage") auf den jeweiligen Wirtsorganismus eingestellt werden. In einer besonders bevorzugten Ausführungsform wird die Eukaryontenzelle mit einem oder mehreren Plasmiden, die eine oder mehrere der Oligonukleotidsequenzen der SEQ. ID. No. 6 - 10 transformiert.

- 10 Für die erfolgreiche Transformation des eukaryontischen Wirts werden die aus gesuchten PC-Gene zunächst in einen Klonierungsvektor kloniert und können anschließend über einen Integrationsvektor in das Genom der Wirtszelle stabil integriert werden. Die Gene können dabei auf einen oder mehrere Klonierungsvektoren verteilt werden. In einer besonders bevorzugten Ausführungsform der Erfindung ist  
15 die Klonierung aller 5 PC-Gene in nur ein Plasmid erfolgt.

Sofern die konstitutive Expression eines Gens auf dem Vektor oder in dem Genom gewünscht ist, kann diesem Gen ein konstitutiver Promotor - beispielsweise ein ADH-Promotor - vorgeschaltet sein. Es sind jedoch auch induzierbare Promotoren,  
20 wie beispielsweise der GAL-Promotor möglich. Die Promotoren können sowohl jedem einzelnen Gen vorgeschaltet sein, als auch die gesamte Gruppe der Gene gemeinschaftlich regulieren.

In einer bevorzugten Ausführungsform wurde ein ADH Promotor gewählt, um die Expression der PC-Gene hox1, cpcE, pcyA und cpcF zu ermöglichen. Der ADH Promotor ist dabei jedem Gen jeweils vorgeschaltet. In dieser Ausführungsform wird das cpcA Gen über den GAL-Promotor reguliert. Damit kann ein induzierbares System für die Biosynthese des Phycocyanins oder des Holo- $\alpha$ -Phycocyanins geschaffen werden.

30 Selbstverständlich ist es alternativ auch möglich, das cpcA-Gen unter die Kontrolle des ADH Promotors und ein oder mehrere der übrigen Gene unter die Kontrolle des GAL Promotors oder eines beliebigen anderen Promotors zu stellen.

35 Sofern ein ADH1-Promotor verwendet wird, ist es von Vorteil, den ADH Promotor nicht in seiner ursprünglichen Länge von etwa 1500 bp zu verwenden, da sich erfahrungsgemäß die Aktivität des ADH1 Promotors über die Zeit reduziert. So konnte beobachtet werden, dass der vollständige Promotor im späten exponen-

tiellen Wachstum der Zellen abschaltet.

Nach Untersuchungen von Ruohonen *et al.* ist diese Eigenschaft des ADH1 Promotors auf ein stromaufwärts gelegenes 700 bp Fragment zurückzuführen. Die Deletion dieser Region hin zu einen "mittleren Promotor" von etwa 700 bp ist dagegen für eine hohe und stabile Expression der klonierten Gene geeignet. Die Publikation von Ruohonen *et al.* wird hinsichtlich der Struktur und Lage eines erfindungsgemäß verwendbaren ADH1 Promotors vollumfänglich in Bezug genommen.

In einer bevorzugten Ausführungsform wird ein ADH1 Promotor mit der Nukleotidsequenz gemäß SEQ. ID. No. 5 eingesetzt.

Da eukaryontische mRNA im Gegensatz zu prokaryontischer mRNA nicht polycistronisch ist, müssen die einklonierten Gene jeweils mit einer Terminationssequenz versehen sein. In einer vorteilhaften Ausführungsform wird die Terminationssequenz nach Guo und Sherman eingesetzt (Guo, Z., Sherman, F. Signals sufficient for 3' end formation of Yeast mRNA. Mol Cell Biol 1996 (16), 2772-2776). Besonders bevorzugt ist die unter Ziffer 3 dargestellte Terminationssequenz SEQ. ID. No. 13.

#### Ausführungsbeispiel

### 1. Klonierungsstrategie

(siehe auch Fig. 1 "Klonierungsstrategie: Aufbau und Klonierung der Gene/Vektoren")

#### 1.1. Klonierung und Synthese der Gene:

Da die Umgebung des Startcodons erhalten bleiben soll, kommen folgende Enzyme in Frage:

Enzym	Erkennungsseq.	Isoschizomere	kompatible Enden	Recut
BspHI	T CATG A A GTAC T		NcoI AflIII (A CATGT)	NlaIII

NcoI	C   CATG G G GTAC   C		AflIII BspHI, PciI	NlaIII
------	--------------------------	--	-----------------------	--------

BspHI und NcoI sind kompatibel und schneiden nicht in den Gensequenzen.

Die Gene wurden über 5' NcoI oder BspHI und 3' XhoI in die Vektoren kloniert. Hierzu erhielten die Gene 5' die jeweils nach Umgebung des Startcodons passende Restriktionsschnittstelle. Am 3'-Ende wurde die Restriktionsschnittstelle für XhoI angefügt. Hier wurde darauf geachtet, dass möglichst wenige Aminosäuren durch die Klonierung zusätzlich in die Sequenz eingefügt wurden. Außerdem durfte durch das Anfügen der XhoI-Schnittstelle kein Frameshift entstehen. Dem wurde durch das Einfügen der Basen AA vor der Schnittstelle und dem Anhängen eines A nach der Schnittstelle Rechnung getragen. Hierdurch wurde an jedes Gen die Aminosäuresequenz NSR angehängt, bei der sichergestellt wurde, dass sie nicht durch seltene Codons codiert wird.

## 1.2. Klonierung und Synthese der Vektoren:

Die Vektoren sollen ein nachfolgendes Ausschneiden des gesamten Gens wahlweise mit und ohne His-Tag erlauben, um einen Nachweis der Expression der Einzelgene zu ermöglichen. Zur Klonierung der Gene hintereinander wurde der HisTag jedoch weggelassen, um eventuelle Probleme mit der zusätzlichen Sequenz zu vermeiden.

Somit enthalten die Vektoren NcoI-XhoI oder BspHI-XhoI je mit und ohne HisTag sowie mit und ohne Terminationssequenz. Durch das vorliegende Design der Vektoren können die Gene in diese einkloniert und danach mit/ohne HisTag, Stopcodon und Terminationssequenz ausgeschnitten und weiter kloniert werden.

Als Ausgangsvektor, in den die synthetisierten mcs und die Gene hintereinander einkloniert wurden, diente pSL1190 (s. Karte von pSL 1180 (Fig. 2); die Vektoren pSL1190 und pSL1180 unterscheiden sich lediglich in der Orientierung der mcs). Die bei der Synthese der vier verschiedenen mcs-Sequenzen berücksichtigten Restriktionsschnittstellen sind in der Figur durch Kästen kenntlich gemacht – so sind mehr Möglichkeiten bei der Klonierung vorhanden. Die letztlich bei der Klonierung der PC-Gene verwendeten Restriktionsschnittstellen sind in der Figur zusätzlich durch einen den jeweiligen Kasten unterlegenden Strich markiert.

Beim Design der mcs für die Vektoren wurde auf folgende Dinge geachtet:

- die verwendeten Enzyme dürfen nicht in den Promotorsequenzen schneiden
- sie dürfen nicht im Rückgrat des Vektors schneiden
- der Abstand zwischen den Restriktionsschnittstellen muss groß genug sein,
- 5 damit benachbarte Enzyme beide schneiden können
- die mcs müssen in den pSL1190 kloniert werden können - hierzu werden die Enzymschnittstellen EcoRI und HindIII an den Enden der mcs vorgesehen
- es müssen Schnittstellen zur Integration der gesamten PC-Kassette (alle 5 Gene hintereinander kloniert) in den Vektor pRS306 vorliegen (hier:
- 10 SacI/Sall)
- In dem Vektor pRS306 muss die Möglichkeit zur Linearisierung im URA3-Gen (Selektionsmarker) bestehen, damit der Vektor integrieren kann und die Kassette genomisch integriert vorliegt.

- 15 Die Sequenzen der mcs der Vektoren sind beigefügt (SEQ. ID. No. 1 - SEQ. ID No. 4).

Aus den Vektoren 656 und 660 (pSL1190 mit synthetisierter mcs, die BspHI-site enthält) müssen, um eine Klonierung mit BspHI durchführen zu können, drei

20 Schnittstellen dieses Enzyms aus dem Rückgrat des Vektors entfernt werden. Hierzu wird der Quik Change Multi Site Directed Mutagenesis Kit von Stratagene eingesetzt und der Erfolg des Experimentes durch Restriktionsanalysen der entstandenen Vektoren verifiziert.

- 25 Entstandene Klone bei der Vektorklonierung: (die weiter verwendeten Klone sind **fett** gekennzeichnet)

Klon-Nr.: B	Klonbeschreibung
647	GeneArt mcs-A = 03-200; NcoI/His-Tag in PCR-Script; <b>647</b>
648	GeneArt mcs-B = 03-201; BspHI/His-Tag in PCR-Script; <b>648</b>
649	GeneArt mcs-C = 03-215; NcoI/ohne His-Tag in PCR-Script; <b>649</b>
650	GeneArt mcs-D = 03-216; BspHI/ohne His-Tag in PCR-Script; <b>650</b>
651 bis 653	GeneArt mcs = 03-199 in pRS306; <b>651</b>
654, 655	GeneArt mcs-A = 03-200 = NcoI/His-Tag in pSL1190, <b>654</b>
656, 657	GeneArt mcs-B = 03-201 = BspHI/His-Tag in pSL1190 <b>656</b>
658, 659	GeneArt mcs-C = 03-215 = NcoI/ohne His-Tag in pSL1190; <b>658</b>
660, 661	GeneArt mcs-D = 03-216 = BspHI/ohne His-Tag in pSL1190; <b>660</b>

682 bis 689	QuikChange mit B656 zur Entfernung BspHI sites; <b>684, 685</b>
690 bis 697	QuikChange mit B660 zur Entfernung BspHI sites; <b>690, 693</b>

### 1.3. Wahl und Klonierung des ADH Promotors:

Der ADH-Promotor (Alkohol-Dehydrogenase) ist in vielen Klonierungsvektoren vorhanden und führt dort zu einer konstitutiven Expression nachgeschalteter Gene. Oftmals ist jedoch bei seiner Verwendung ein Rückgang der Expressionsrate des klonierten Gens bis hin zu einem kompletten Abschalten der Expression nach etwa 3 Tagen festzustellen.

Es hat sich aber gezeigt, dass ein ein 700 bp langes Stück des originären (kompletten) Promotors umfassender ADH Promotor zu einer Expression der klonierten Gene während des gesamten Wachstumszyklus der Hefe führt (Ruohonen et al. (1995) a.a.O.).

Dieses Stück des Promotors war der Ausgangspunkt für den erfindungsgemäß verwendeten ADH-Promotor, wobei allerdings bei der Klonierung des Promotors die Primer für die PCR so gewählt wurden, dass ein ADH-Fragment von insgesamt 719 bp entstand (siehe SEQ. ID. No. 5).

Die Synthese des ADH Promotors wurde mittels PCR aus dem Genom von *S. cerevisiae* Y190 durchgeführt. Die Klonierung erfolgte in einer Drei-Fragmente-Ligation mit dem fraglichen Gen und dem Vektor zusammen. Hierfür wurde immer der gleiche reverse-Primer eingesetzt, der eine NcoI-Schnittstelle besitzt, da NcoI und BspHI kompatibel sind. Der forward-Primer besitzt jeweils eine andere angehängte Restriktionsschnittstelle, die je nach Klonierungsschritt ausgewählt wird.

Sequenzen der verwendeten Primer (die Restriktionsschnittstelle ist jeweils unterstrichen): Die zusätzlichen Basen werden benötigt, damit das Enzym das PCR-Fragment schneiden kann:

ADH-SacI-for	CAATTAGAGCTCATATCCTTTTGTGTTTCCGGGTG
ADH-SacII-for	ATCCCCGCGGATATCCTTTTGTGTTTCCGGGTG
ADH-NotI-for	AAGGAAAAAAGCGGCCGCATATCCTTTTGTGTTTCCGGGTG



ADH-XbaI-for	AGCTCTAGAATATCCTTTTGTTGTTTCCGGGTG
ADH-BglII-for	CGAAGATCTATATCCTTTTGTTGTTTCCGGGTG
ADH-NcoI-rev	GTAAGCCATGGTGTATATGAGATAGTTGATTGTATGCTT GG

## 2. Optimierung der Codon-Usage

Die Codon-Usage wurde für die Häufigkeit der Codonnutzung (Codon-Usage) von *S. cerevisiae* optimiert. Der GC-Gehalt wurde, wenn möglich, auf einen durchschnittlichen Wert von ca. 45 % eingestellt, um die genetische Stabilität der mRNA-Halbwertszeit zu steigern.

Darüber hinaus wurden folgende für die Klonierung vorgesehenen Restriktionschnittstellen aus den Genen entfernt:

Gen	Restriktionsenzym	Position (bp)
CpcA	KpnI	312
CpcE	NheI	779
	YbaI	493
	HpaI	366
	HpaI	549

Es wurden für die Klonierung der für die Expression des Holo-CpcA erforderlichen Gene ("PC-Gene") folgende Sequenzen verwendet:

Gen	SEQ. ID. No.
cpcA	SEQ. ID. No. 6
cpcE	SEQ. ID. No. 7
cpcF	SEQ. ID. No. 8
hox1	SEQ. ID. No. 9
pPcyA	SEQ. ID. No. 10
cpcB	SEQ. ID. No. 11

Sofern ein Holo-CpcB exprimiert werden soll, kann alternativ zum cpcA das cpcB-Gen einkloniert werden (SEQ. ID. No. 11). Das codiert eine Aminosäuresequenz gemäß SEQ. ID. No. 12.

## 3. Wahl der Terminationssequenz

Da eukaryotische mRNAs im Gegensatz zu prokaryontischen nicht polycistronisch sind, muß eine Termination nach jedem Gen stattfinden. Guo und Sherman (1996, a.a.O.) beschreiben eine synthetische DNA-Sequenz, die für die Bildung eines 3'-Endes ausreicht und aus "efficiency element", "positioning element" und "poly(A) region" besteht.

Für unsere Zwecke wurde die XbaI-Schnittstelle entfernt und die Spacer-Region zwischen "efficiency element" und "positioning element" wie folgt gewählt: ACTCTGTAGA (statt ACTGTCTAGA; XbaI-Schnittstelle (SEQ. ID. No. 13)).

#### 4. Klonierung der Einzelgene und Test der Expression der Einzelgene

Das Vorgehen bei der Klonierung der Einzelgene sowie für den Nachweis ihrer erfolgreichen Expression ist schematisch in Fig. 3 dargestellt. Die Darstellung bezieht sich exemplarisch auf den Klon B654:pSL1190 mit NcoI-Schnittstelle und HisTag.

Die Linearisierung des fertigen Vektors erfolgte mit StuI. Die Transformation erfolgte in den Hefestamm Y190 mit URA Defizienz.

Zunächst wurden Vortests mit zwei anderen Fluorophoren durchgeführt, um zu testen, ob eine Expression der Gene hinter dem gewählten ADH-Promotor und mit der eingesetzten Terminationssequenz erfolgt.

Für diesen Nachweis wurden die Gene einzeln jeweils mit dem ADH-Promotor in den Vektor pRS306 kloniert. Dabei handelt es sich um einen Integrationsvektor, so dass die Gene anschließend in das Genom der Hefe integriert werden können. Damit entspricht die Situation im wesentlichen dem angestrebten Ziel, nämlich dass die fünf für das Holo-CpcA relevanten Gene ("PC-Gene") hintereinander liegen. Als Schnittstellen zur Integration des Gens mit Promotor in den Vektor pRS306 wurde SacI/SalI gewählt, da so am meisten Fremdsequenz in die Hefe eingebracht werden kann.

Hierzu wurden die Gene zuerst in den passenden Vektor (pSL1190 mit BspHI oder NcoI-site und HisTag) kloniert, um anschließend in einer Drei-Fragmente Ligation mit dem ADH-Promotor ligiert zu werden.

Zur Durchführung der Drei-Fragmente-Ligation wurden die eingesetzten DNA-Mengen mittels Agarosegel mit mitgeführtem Marker bestimmt. Danach wurden

die eingesetzten DNA-Mengen auf molarer Ebene ausgerechnet. Dem Vektor wurden 25 fmol und den beiden Inserts je 125 fmol eingesetzt.

Das Zellpellet nach der Kultivierung von Protein-exprimierenden Hefen wurde mit *glass beads* und Vortexer aufgearbeitet. Hierbei wurden denaturierende Bedingungen gewählt. Bei der Aufreinigung des HisTag tragenden Proteins wurden die Vorschriften des Herstellers der Ni-NTA-Säulchen eingehalten.

### 5. Klonierung der Gene für das Holo-CpcA hintereinander

Das Vorgehen für die Klonierung der PC-Gene hintereinander ist schematisch in der Fig. 4a bis 4c dargestellt. Die Karte des Vektors mit den einklonierten PC-Genen zeigt Fig. 5. Die Sequenzdaten für den Vektor enthält SEQ. ID. No. 14. Sofern dem CpcA der Gal-Promotor vorgeschaltet ist, zeigt der Vektor einen Aufbau nach Fig. 6.

Die Gene wurden in pSL 1190 mit synthetisierter mcs ohne HisTag kloniert. Danach wurden diese Gene mit passenden Enzymen herausgeschnitten und zusammen mit dem ADH-Promotor hintereinander in den Vektor pSL1190 gesetzt. Dabei wurde der Erfolg der Klonierungen mittels Bakterienkolonie-PCR für das jeweilige Gen nachgewiesen und die Identität der isolierten DNA wurde durch Testcuts verifiziert. Zusätzlich wurde der Promotoranteil der jeweils entstandenen Vektoren sequenziert (AGOWA), da dieser durch PCR erzeugt wurde (variierender forward primer mit verschiedenen Restriktionsenzymen) und Fehler vorhanden sein könnten. Für diese PCR-Anwendungen wurde die proof-reading-Polymerase Pfu (Promega) eingesetzt.

Nachdem die Gene hintereinander in dem Vektor pSL1190 vorlagen, wurde die gesamte Kassette über die Restriktionsschnittstellen SacI/Sall in den Integrationsvektor pRS306 kloniert. Bei diesem Klonierungsschritt wurden das cpcA-Gen mit passenden Primern und der Vektor pRS306 mittels einer PCR mit Primern, die ein Stück aus dem URA3-Selektionsmarker amplifizieren, in den transformierten Bakterienkolonien nachgewiesen.

Sequenz der zum Nachweis der klonierten Gene/Integrationsvektor pRS306 verwendeten Primer:

		Länge des Fragments
--	--	---------------------

cpcF-for	TGATTCAAGCTGTTGAAACC	361 bp
cpcF-rev	GATTGCCATCTCAAGAAACC	
hox-for	CTGCTGGTCAAGCATACG	413 bp
hox-rev	GGTAGCCAAACCAACTTCG	
cpcA-for	CTTTCGGTAGATTGAGACAAGC	414 bp
cpcA-rec	CAAAGCGTTAATAGCGTAATCC	
pcyA-rev	TTGACCAACTCTTCTTTGATGC	579 bp
pcyA-rev	AGCTTCAGACAATGGTTCAGC	
cpcE-for	AAGATGAAACCGATAGATCACC	609 bp
cpcE-rev	AAGCCTTAGATTCAGAAGATGC	
URA-pRS-for	GAACGTGCTGCTACTCATCC	613 bp
URA-pRS-rev	CGTCTCCCTTGTCATCTAAACC	

Um das Holo-Protein mittels Western-Blot und Immunodetektion gegen den HisTag nachweisen zu können, wurde das cpcA-Gen mit HisTag hinter die Kasse-  
 5 tte mit den restlichen Genen gesetzt und dann diese vollständige Kassette in pRS306 kloniert. Nach der Transformation mit nachfolgender Expression des Proteins in der Hefe wurde das Protein aufgereinigt und in einem Western Blot dargestellt (Fig. 7).

#### Western Blot:

10 Für den Western Blot wurde als primärer Antikörper der *Mouse Anti Histidine Tag* (Serotec) 1:250 eingesetzt. Als sekundärer Antikörper diente *Goat Anti Mouse Alkaline Phosphatase* (Dianova) 1:5000. Die Detektion erfolgte mit BCIP/NBT-Blue Liquid (Sigma).

15 Als Marker wurde der prestained SDS-Gel Marker (Invitrogen) verwendet.

#### **6. Spektrum des erfindungsgemäßen Holo-CpcA**

20 Das Spektrum (Fig. 8) des erfindungsgemäßen Holo-CpcA wurde in folgendem Puffer aufgenommen:

#### Lysepuffer:

50 mM NaH<sub>2</sub>PO<sub>4</sub>  
 300 mM NaCl  
 25 10 mM Imidazol

pH 8,0

Elutionspuffer:

50 mM NaH<sub>2</sub>PO<sub>4</sub>

5 300 mM NaCl

250 mM Imidazol

pH 8,0

Der Lysepuffer und der Elutionspuffer lagen im Verhältnis 60% zu 40% vor.

10

### 7. Hefestamm und Kultivierungsbedingungen

Als Wirtsorganismus für die erfindungsgemäße heterologe Expression der PC-Gene wurde eine Uracil defiziente Variante des Y190 Stamms der Hefe verwendet. Der Hefestamm Y 190 zeigte ursprünglich folgenden Genotyp: MATa, leu2-3,  
15 leu2-112, ura3-52, trp1-901, his3-Δ200, ade2-101, GAL4-Δgal80-Δ, URA3::GAL-LacZ, LYS::GAL-HIS3, cyh<sup>r</sup>.

Zur Bereitstellung einer Uracil defizienten Y190 Hefe wurden aus diesem Stamm nach der Methode von Boeke *et al.* [Boeke, J., Lacroute, F. and Fink, G. R (1984).

20

A positive selection for mutants lacking orotidine-5-phosphate decarboxylase activity in yeast: 5-Fluoro-orotic acid resistance. Mol. Gen. Genet 197, 345-346] Hefezellen selektiert, die eine Deletion des URA3-Gens aufweisen.

Diese URA3 Deletionsmutanten wurden auf YPAD-Medium oder auf beliebigen  
25 synthetischen Komplett-Medien kultiviert. Es hat sich als besonders vorteilhaft erwiesen, anstelle der in der Literatur vorgeschlagenen 2 %-igen Glucose-Konzentration eine bis zu 13 %-ige Glucosekonzentration zu verwenden. Es sind aber alle Konzentrationen zwischen 2 und 20 % möglich.

**Patentansprüche:**

- 5 1. Rekombinante Eukaryontenzelle synthetisierend funktionell aktives Phycocyanin oder funktionell aktive Teile daraus.
2. Rekombinante Eukaryontenzelle synthetisierend Holo-CpcA und/oder Holo-CpcB.
- 10 3. Rekombinante Eukaryontenzelle nach Anspruch 1 oder 2 exprimierend die heterodimere PCB Lyase sowie das Apo-CpcA und/oder das Apo-CpcB.
4. Rekombinante Eukaryontenzelle nach Anspruch 3 exprimierend des weiteren eine heterologe Hämogxygenase und/oder eine heterologe Ferredoxin Oxidoreduktase.  
15
5. Rekombinante Eukaryontenzelle nach einem der Ansprüche 3 oder 4, **gekennzeichnet durch** eines oder mehrere der Gene hox 1, pcyA, cpcE, cpcF und cpcA.  
20
6. Rekombinante Eukaryontenzelle nach Anspruch 5 transformiert mit einem Vektor umfassend mindestens ein Insert mit einer der folgenden Oligonukleotidsequenzen gemäß einer der SEQ. ID. No. 6 - SEQ. ID. No. 10 oder mit einer dieser Sequenzen komplementären Strang hybridisierenden Oligonukleotiden.  
25
7. Rekombinante Eukaryontenzelle nach einem der Ansprüche 1 bis 6, **dadurch gekennzeichnet**, dass mindestens eines der Gene von einem ADH Promotor reguliert wird, der upstream von -700 bp deletiert ist.  
30
8. Rekombinante Eukaryontenzelle nach Anspruch 7, **dadurch gekennzeichnet**, dass alle Gene von einem ADH Promotor reguliert werden, der upstream von -700 bp deletiert ist.
- 35 9. Rekombinante Eukaryontenzelle nach Anspruch 7 oder 8, **gekennzeichnet durch** einen ADH Promotor mit der Oligonukleotidsequenz gemäß SEQ. ID. No. 5 oder mit einer mit dem dazu komplementären Strang hybridisierenden

Oligonukleotidsequenz.

10. Rekombinante Eukaryontenzelle nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass sie eine Hefezelle ist.
11. Hefezelle mit der DSMZ Aufnahme­nummer DSM 16134.
12. Vektor umfassend Inserts codierend für eine Hämoxygenase, eine Ferredoxin Oxidoreduktase, eine heterodimäre PCB Lyase und das Apo-CpcA oder Apo-CpcB.
13. Vektor nach Anspruch 12, **gekennzeichnet durch** Inserts mit den Genen hox1, pcyA, cpcE, cpcF und cpcA, wobei das cpcA auch durch cpcB ersetzt werden kann.
14. Vektor nach Anspruch 11, **gekennzeichnet durch** eine oder mehrere der Oligonukleotide gemäß einer der SEQ. ID. No. 6 - SEQ. ID. No. 10 oder mit einer dieser Sequenzen komplementären Strang hybridisierenden Oligonukleotiden.
15. Vektor nach einem der Ansprüche 12 bis 14, **dadurch gekennzeichnet**, dass mindestens eines der Gene von einem ADH Promotor reguliert wird, der upstream von -700 bp deletiert ist.
16. Vektor nach Anspruch 15, **dadurch gekennzeichnet**, dass alle Gene von einem ADH Promotor reguliert werden, der upstream von -700 bp deletiert ist.
17. Vektor nach Anspruch 15 oder 16, gekennzeichnet durch einen ADH Promotor mit der Oligonukleotidsequenz gemäß SEQ. ID. No. 5 oder mit einer mit dem dazu komplementären Strang hybridisierenden Oligonukleotidsequenz.
18. Vektor mit einer Nukleotidsequenz gemäß SEQ. ID. No. 14.
19. Verfahren zur Herstellung eines Phycocyanins oder eines funktionell aktiven Teils davon, insbesondere Holo-CpcA oder Holo-CpcB umfassend folgende Schritte:

- Transformation einer eukaryontischen Wirtszelle mindestens mit cpcA oder cpcB;
  - ggf. Addition eines Phycocyanobilins (PCB) und/oder der PCB Lyase
  - Isolierung des Phycocyanins, des Holo-CpcA und/oder des Holo-CpcB aus der Wirtszelle.
20. Verfahren nach Anspruch 19, **dadurch gekennzeichnet**, dass zusätzlich auch eines oder mehrere der Gene hox1, pcyA, cpcE und/oder cpcF rekombinant exprimiert werden.
21. Verfahren nach einem der Ansprüche 19 oder 20, **dadurch gekennzeichnet**, dass die endogene Hämkonzentration in der Wirtszelle erhöht wird.
22. Verfahren nach Anspruch 19 bis 21, **gekennzeichnet durch** die Stimulierung des endogenen Hämstoffwechsels in der Wirtszelle.
23. Verfahren nach Anspruch 22, **gekennzeichnet durch** die Verwendung einer nicht von der Wirtszelle fermentierbaren Kohlenstoffquelle.
24. Verfahren nach einem der Ansprüche 19 bis 21, **gekennzeichnet durch** eine Inhibierung des endogenen Hämabbaus in der Wirtszelle.
25. Verfahren nach einem der Ansprüche 19 bis 24, **gekennzeichnet durch** die Erhöhung der Glukosekonzentration im Medium der Wirtszelle auf eine Konzentration, die die Aktivität des ADH-Promotors stimuliert.
26. Verfahren nach Anspruch 25, **gekennzeichnet durch** die Gabe von mindestens 13 % Glucose.
27. Verfahren nach einem der Ansprüche 19 bis 26, **gekennzeichnet durch** eine 2 %-ige Ethanolkonzentration im Kulturmedium der Wirtszelle.
28. Verfahren nach einem der Ansprüche 19 bis 27, **gekennzeichnet durch** eine Wirtszelle nach einem der Ansprüche 1 bis 11.



29. Verfahren nach einem der Ansprüche 19 bis 26, **dadurch gekennzeichnet**, dass die Transformation mit Hilfe eines Vektors nach einem der Ansprüche 12 bis 19 erfolgt.

5 30. Rekombinantes Phycocyanin oder funktionell aktive Teile daraus, hergestellt nach einem der Ansprüche 19 bis 29.

31. Rekombinantes Phycocyanin oder funktionell aktive Teile daraus nach Anspruch 30, **gekennzeichnet durch** eine Glykosylierung.

10

32. Rekombinantes Phycocyanin oder funktionell aktive Teile daraus nach einem der Ansprüche 30 oder 31, **dadurch gekennzeichnet**, dass die Glykosylierung an mindestens einer der Position 6, 10 oder 162 der Aminosäuresequenz vorliegt.

15

33. Rekombinantes Phycocyanin oder funktionell aktive Teile daraus nach einem der Ansprüche 30 bis 32, **dadurch gekennzeichnet**, dass das Emissionsmaximum im Vergleich zum nativen Phycocyanin, insbesondere zum native Holo-CpcA bei pH 8 in den längerwelligen Bereich verschoben ist.

20

34. Rekombinantes Holo-CpcA mit einem Emissionsmaximum bei pH 8 von 643 nm.

25

35. Fusionsprotein enthaltend ein rekombinantes Phycocyanin oder funktionell aktive Teile davon, insbesondere rekombinantes Holo-CpcA, nach einem der Ansprüche 30 bis 34.

30

36. Antikörper oder Fragmente davon, **dadurch gekennzeichnet**, dass er spezifisch ein rekombinantes Phycocyanin nach einem der Ansprüche 30 bis 34 erkennt.

37. Antikörper oder Fragmente davon, **dadurch gekennzeichnet**, dass er spezifisch funktionell aktive Teile des rekombinantes Phycocyanins nach einem der Ansprüche 30 bis 34, insbesondere rekombinantes Holo-CpcA, erkennt.

Fig. 1

# Klonierungsstrategie – Aufbau und Klonierung der Gene / Vektoren

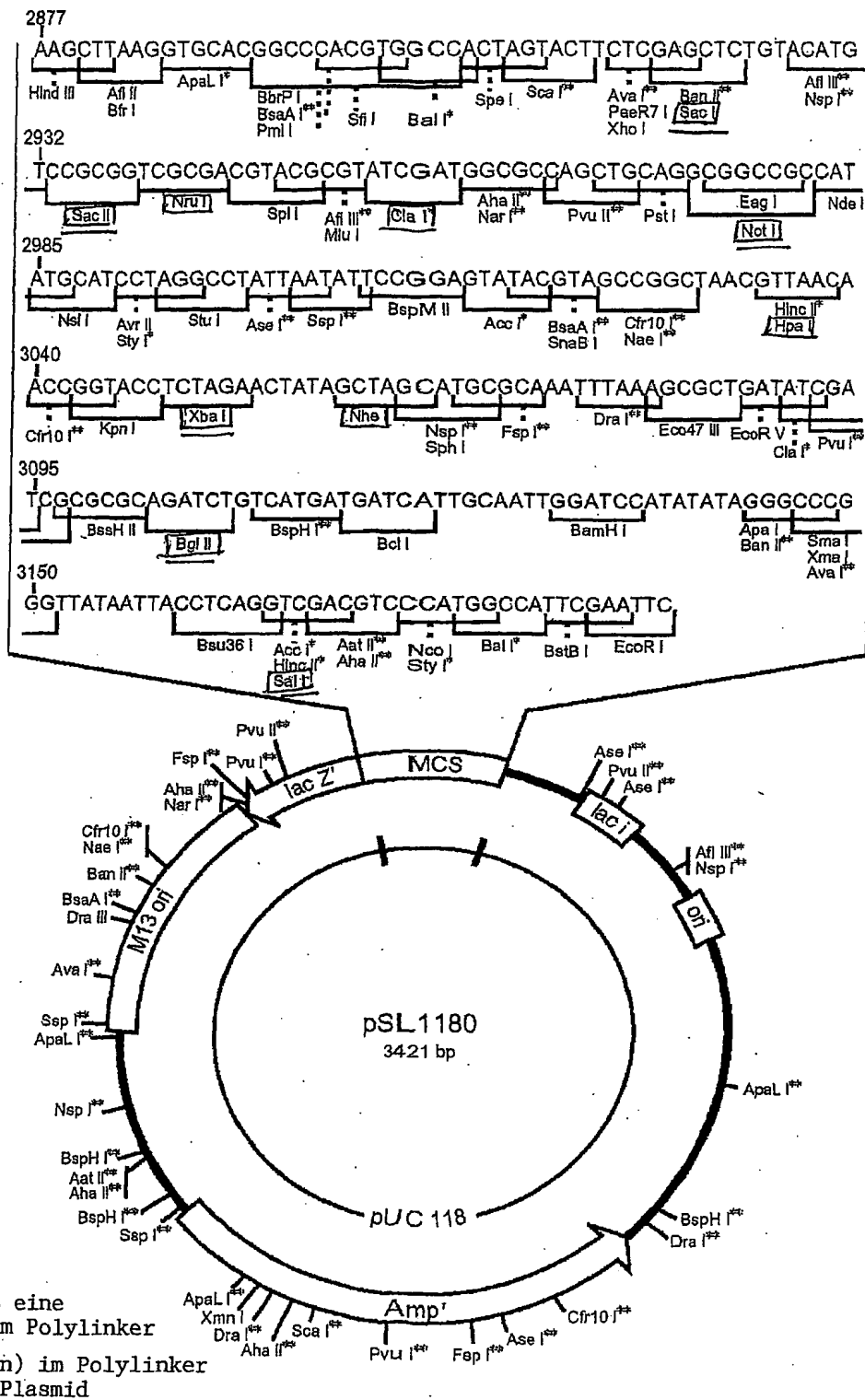
Gene: PC-Gen NcoI: CCATGG  
 NcoI oder BspHI XhoI BspHI: TCATGA

Vektoren:  $\text{H}^{\text{NcoI-XhoI-}} \text{His} \text{-TAA-} \text{E} \text{-S-N-X-B-Sal}$   
 $\text{H}^{\text{NcoI-XhoI- TAA-}} \text{T} \text{-S-N-X-B-Sal}^{\text{E}}$   
 $\text{H}^{\text{BspHII-XhoI-}} \text{His} \text{-TAA-} \text{T} \text{-S-N-X-B-Sal}^{\text{E}}$   
 $\text{H}^{\text{BspHII-XhoI- TAA-}} \text{T} \text{-S-N-X-B-Sal}^{\text{E}}$

S: SacI  
 N: NotI  
 X: XbaI  
 B: BglII  
 His: HisTag  
 T: Terminations-  
 sequenz

Fig. 2

2/10



ERSATZBLATT (REGEL 26)

Fig. 3

# Vorgehen bei der Klonierung – Test der Expression der Einzelgene

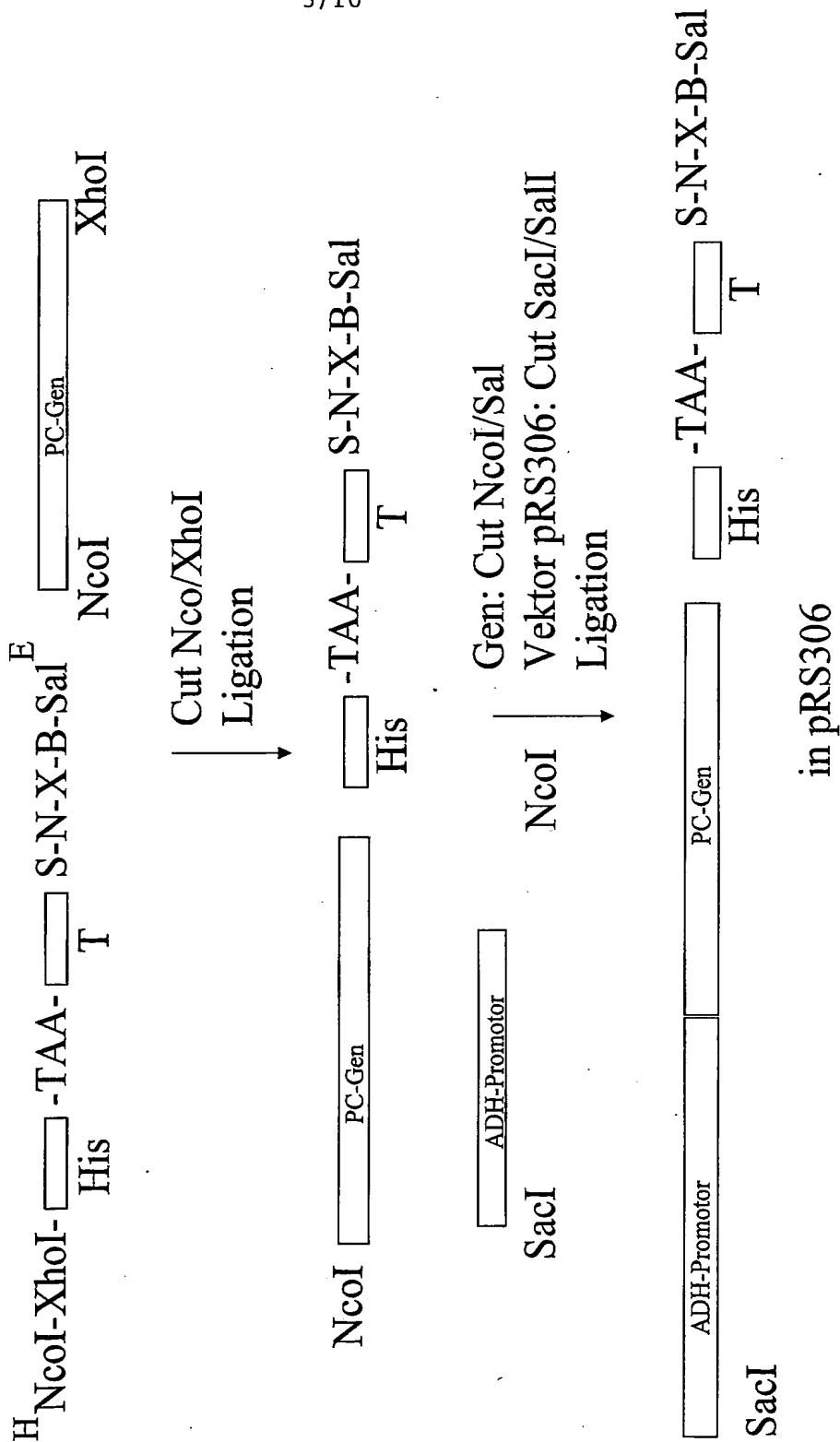


Fig. 4 a

# Vorgehen bei der Klonierung – Klonierung der Gene hintereinander

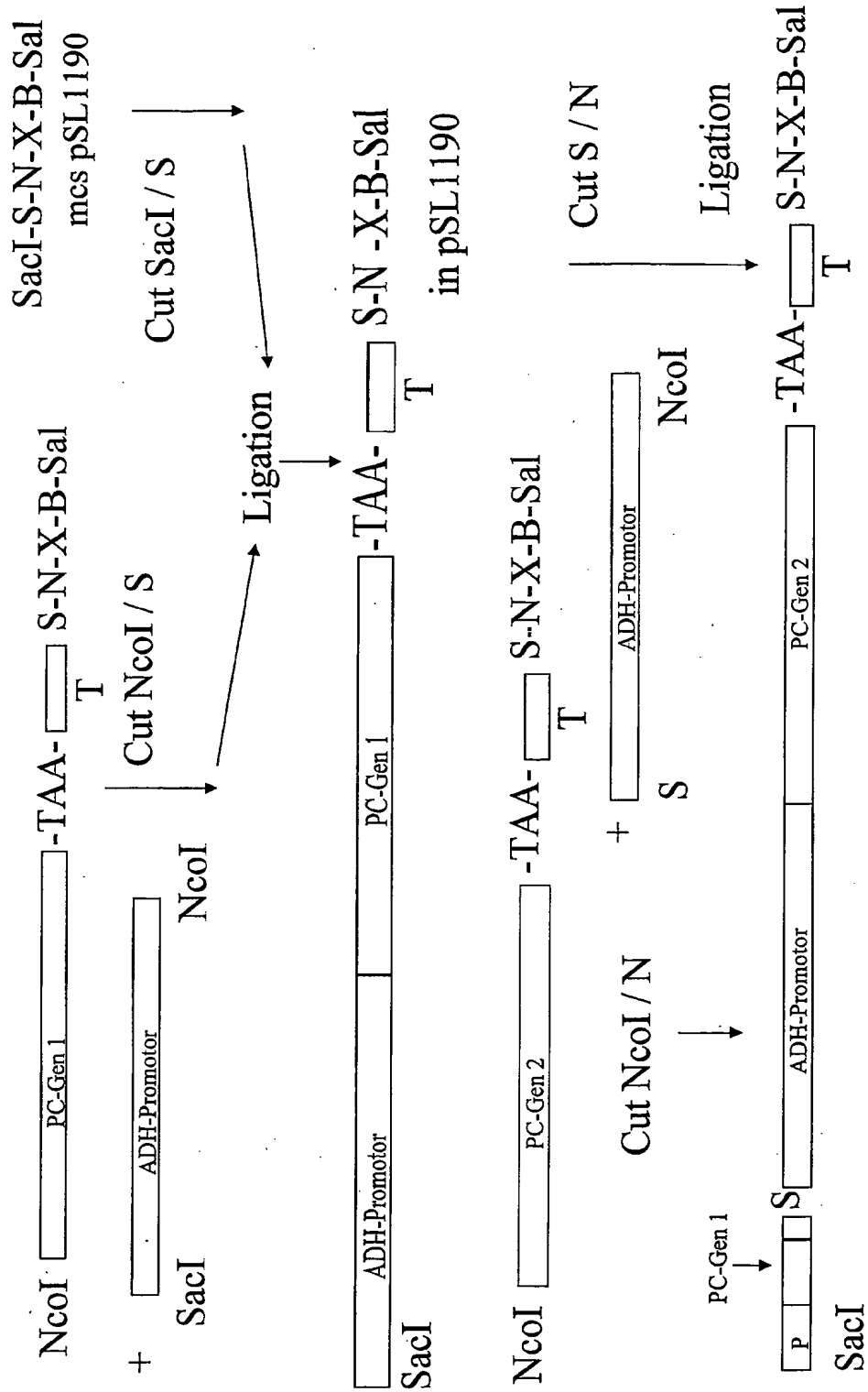
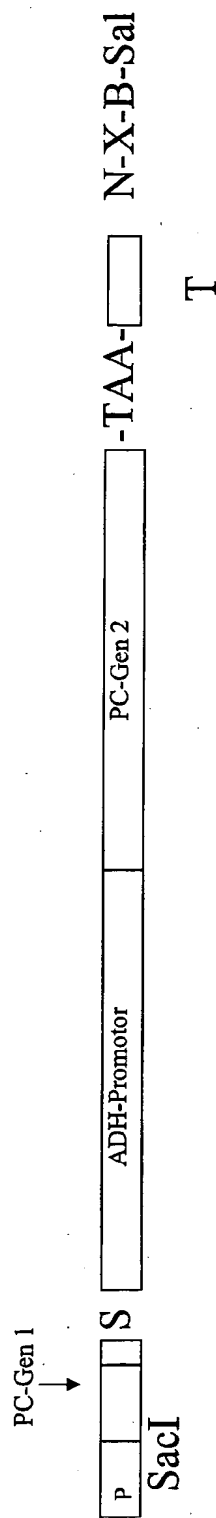


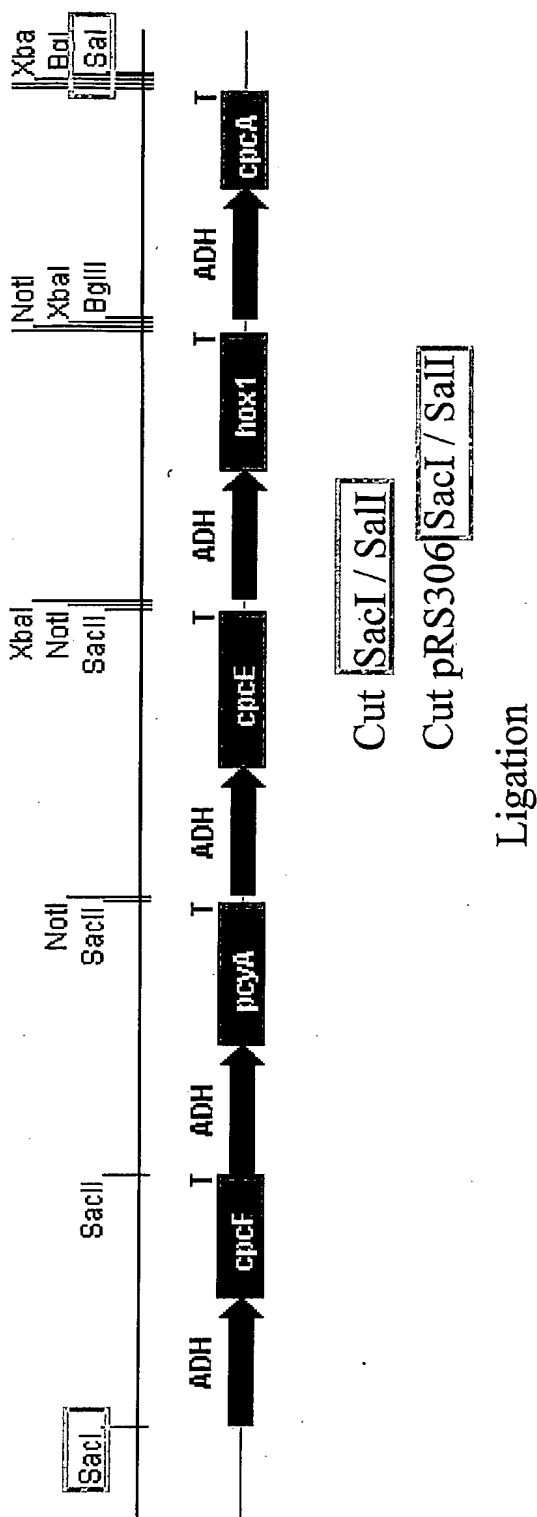
Fig. 4 b

Vorgehen bei der Klonierung – Klonierung der Gene hintereinander



usw. – jetzt Cut mit N / X

ERSATZBLATT (REGEL 26)



Cut SacI / SalI

Cut pRS306 SacI / SalI

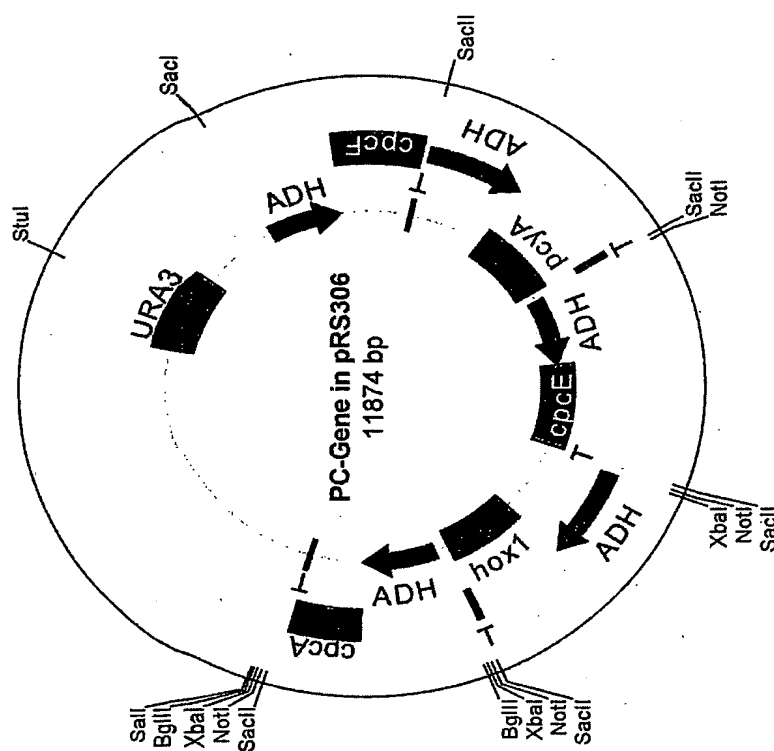
Ligation

**Fig. 4c**

- 1) ADH-Promotor (SacI/ NcoI, 1) plus cpcF aus 706 (NcoI/SacII, 4) in pSL1190 (SacI/ SacII, 4)
- 2) ADH-Promotor (SacII/ NcoI, 4) plus pcyA aus 711 (NcoI/NotI, 3) in pSL1190 (SacII/ NotI, 4) mit cpcF
- 3) ADH-Promotor (NotI/ NcoI, 3) plus cpcE aus 714 (BspHI/ XbaI, 2) in pSL1190 (NotI/ XbaI, 3) mit cpcF und pcyA
- 4) ADH-Promotor (XbaI/NcoI, 2) plus hox1 aus 808 (BspHI/ BglII, seq) in pSL1190 (XbaI/ BglII, 2) mit cpcF und pcyA und cpcE
- 5) ADH-Promotor (BglII/NcoI, 3) plus cpcA aus 703 (BspHI/ Sall, seq) in pSL1190 (BglII/ Sall, seq) mit cpcF und pcyA und cpcE und hox1  
Gesamte Kassette aus pSL1190 (SacI/Sall, seq) in pRS306 (SacI/ Sall, seq) → Test der konstitutiven PC-Expression
- 6) GAL-Promotor (BglII/NcoI, 3) plus cpcA aus 703 (BspHI/ Sall, seq) in pSL1190 (BglII/ Sall, seq) mit cpcF und pcyA und cpcE und hox1  
Gesamte Kassette aus pSL1190 (SacI/Sall, seq) in pRS306 (SacI/ Sall, seq) → Test der regulierbaren PC-Expression

7/10

Fig. 5





**Fig. 6**

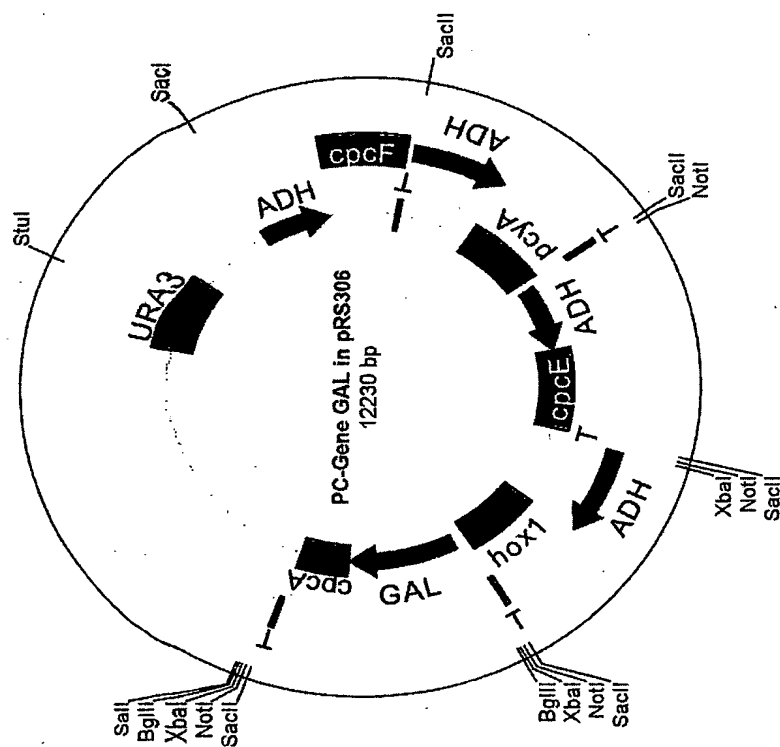


Fig. 7

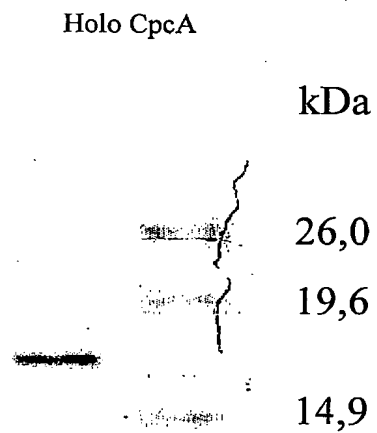
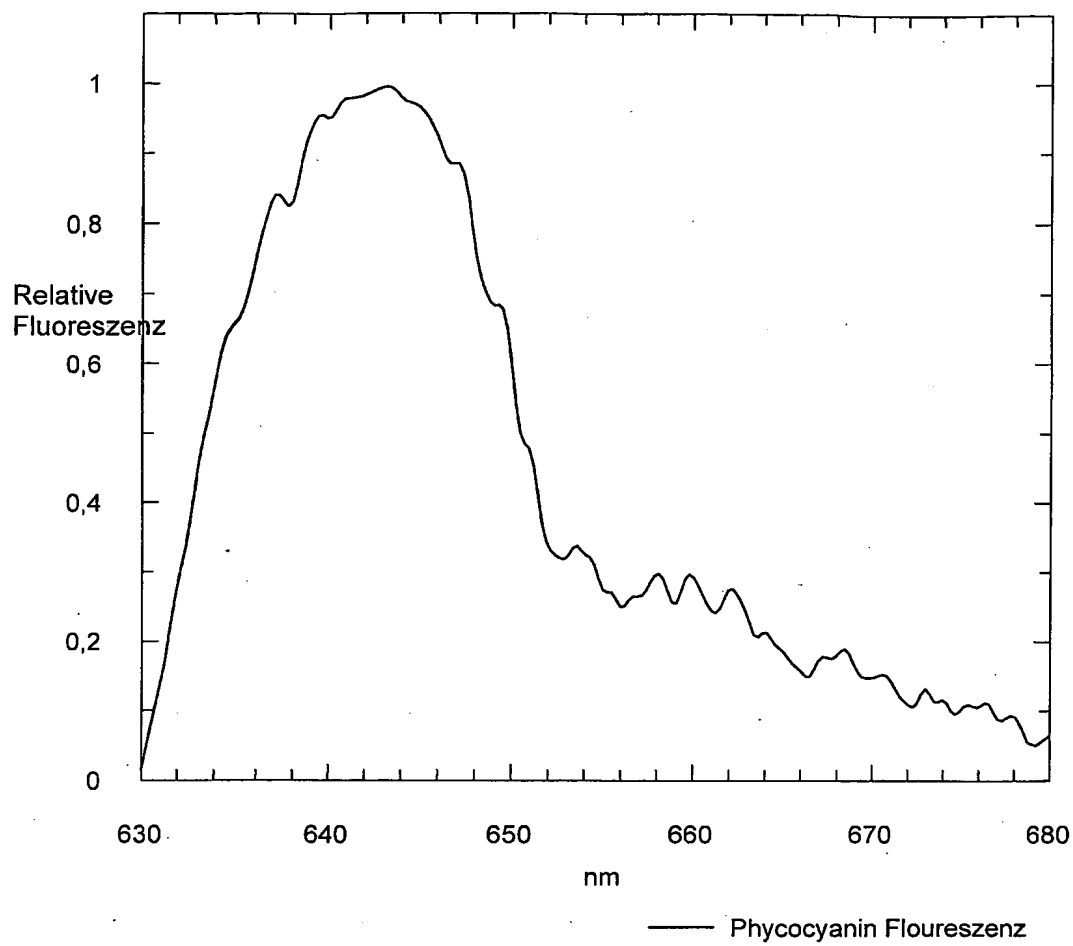


Fig. 8



SEQ. ID. NO 1

NheI BglII Sall EcoRI

ATT GGGCTAGCGTTGAGATCTTTAGAAACGTCGACGAATTC  
TAA CCGGATCGCAACTCTAGAAATCTTTGCAGCTGCTTAAG

160b 180b

SEQ. ID No 2

NheI BglII Sall EcoRI

ATTGGGCTAGCGTTGAGATCTTTAGAAACGTCGACGAATTC  
TAACCCGATCGCAACTCTAGAAATCCTTGCAGCTGCTTAAG

160b 180b

SEQ. ID No 3

dIII                      NcoI                      XhoI                      Stop  
 AAGCTTCATGGAATCTCGAGATAAGGTATATAACTCTGTAGAAATAAAG  
 TTCGAGAGGTACCTTAGAGCTTCTATTCCATATATTGAGACATCTTTATTTC

SacII
NruI
ClaI

AGTATCATCTTTCAAACCGCGGATTGTCGCGATCAAATCGATATGTCTTA  
 TCATAGTAGAAAGTTTGGCGCCTAACAGCGCTAGTTTAGCTATACAGAAT

NotI HpaI XbaI NheI BglII

TGCGGCCGCTTACGACCGTTAACTTGTTCCTAGATTGGGCTAGCGTTGAGA  
ACGCCGCGCAATGCTGGCAATTGAACAAGATCTAACCCGATCGCAACTCT

SalI                      EcoRI

TCTTTAGAAACGTCGACGAATTC  
AGAAATCTTGCAGCTGCTTAAG

03-216  
173 bp

SEA. ID 4

HindIII BspHI XhoI *Stop*  
AAGCTTT CATGAAATCTCGAGATAAGGTATATAACTCTGTAGAAATAAAG  
TTCGAAAGTACTTTAGAGCTCTATTCCATATATTGAGACATCTTTATTTCTC  
0b 20b 40b

SacII NruI ClaI  
AGTATCATCTTTCAAACCGCGGATTGTCTGCGATCAAATCGATATGTCCTTA  
TCATAGTAGAAAGTTTGGCGCCTAACAGCGCTAGTTTAGCTATACAGAAT  
60b 80b 100b

NotI HpaI XbaI NheI BglII  
TGCGGCCGCTTACGACCGTTAACTTGTTCTAGATTGGGCTAGCGTTGAGA  
ACGCCGGCGAATGCTGGCAATTGAACAAGATCTAACCCGATCGCAACTCT  
100b 120b 140b

Sall EcoRI  
TCTTTAGAAACGTCGACGAATTC  
AGAAATCTTTGCAGCTGCTTAAG  
160b

Seq. ID. 5

*AOH - primus - rev*

GAGCTCATAT	CCTTTTGTG	TTCCGGGTG	TACAATATGG	ACTTCCTCTT	50
CTCGAGTATA	GGAAAACAAC	AAAGGCCAC	ATGTTA TACC	TGAAGGAGAA	
TTCTGGCAAC	CAAACCCATA	CATCGGGATT	CCTATAATAC	CTTCGTTGGT	100
AAGACCGTTG	GTTTGGGTAT	GTAGCCCTAA	GGATAT TATG	GAAGCAACCA	
CTCCCTAACA	TGTAGGTGGC	GGAGGGGAGA	TATACAATAG	AACAGATACC	150
GAGGGATTGT	ACATCCACCG	CCTCCCTCT	ATATGT TATC	TTGTCTATGG	
AGACAAGACA	TAATGGGCTA	AACAAGACTA	CACCAATTAC	ACTGCCTCAT	200
TCTGTTCTGT	ATTACCCGAT	TTGTTCTGAT	GTGGTTAATG	TGACGGAGTA	
TGATGGTGGT	ACATAACGAA	CTAATACTGT	AGCCCTAGAC	TTGATAGCCA	250
ACTACCACCA	TGTATTGCTT	GATTATGACA	TCGGGATCTG	AACTATCGGT	
TCATCATATC	GAAGTTTCAC	TACCCTTTTT	CCATTTCGCA	TCTATTGAAG	300
AGTAGTATAG	CTTCAAAGTG	ATGGGAAAAA	GGTAAACGGT	AGATAACTTC	
TAATAATAGG	CGCATGCAAC	TTCTTTTCTT	TTTTTTTCTT	TTCTCTCTCC	350
ATTATTATCC	GCGTACGTTG	AAGAAAAGAA	AAAAAAAGAA	AAGAGAGAGG	
CCCGTTGTTG	TCTCACCATA	TCCGCAATGA	CAAAAAGAA	ATGGAAGACA	400
GGGCAACAAC	AGAGTGGTAT	AGGCGTTACT	GTTTTTTTAC	TACCTTCTGT	
CTAAAGGAAA	AAATTAACGA	CAAAGACAGC	ACCAACAGAT	GTCGTTGTTC	450
GATTTCTCTT	TTTAATTGCT	GTTTCTGTCG	TGGTTCTCTA	CAGCAACAAG	
CAGAGCTGAT	GAGGGGTATC	TCGAAGCACA	CGAAACTTTT	TCCTTCCTTC	500
GTCTCGACTA	CTCCCCATAG	AGCTTCGTGT	GCTTTGAAAA	AGGAAGGAAG	
ATTCACGCAC	ACTACTCTCT	AATGAGCAAC	GGTATACGGC	CTTCCTTCCA	550
TAAGTGGGTG	TGATGAGAGA	TTACTCGTTG	CCATA TGCCG	GAAGGAAGGT	
GTTACTTGAA	TTTGAAATAA	AAAAAAGTTT	GCTGT CTGTC	TATCAAGTAT	600
CAATGAACCT	AAACTTTATT	TTTTTTTCAA	CGACA GAACG	ATAGTTCATA	
AAATAGACCT	GCAATTATTA	ATCTTTTGTT	TCCTC GTCAT	TGTTCTCGTT	650
TTTATCTGGA	CGTTAATAAT	TAGAAAACAA	AGGAG CAGTA	ACAAGAGCAA	
CCCTTTCTTC	CTTGTTTCTT	TTTCTGCACA	ATATT TCAAG	CTATACCAAG	700
GGGAAAGAAG	GAACAAAGAA	AAAGACGTGT	TATAAAGTTC	GATATGGTTC	
CATACAATCA	ACTATCTCAT	ATACACCATG	G		750
GATATGTTAGT	TGATAGAGTA	TATGTGGTAC	C		

*AOH primus rev*



Seq. ID. 6

~~Seq. ID. 6~~  
 TCATGAAGAC CCCATTGACC GAAGCTGTTT CTACCGCAGA TTCTCAAGGT 50  
 AGTACTTCTG GGGTAACTGG CTTGACACAA GATGGCGTCT AAGAGTTCCA  
 AGATTCTTGT CATCTACCGA ATTGCAAATT GCTTTCGGTA GATTGAGACA 100  
 TCTAAGAACA GTAGATGGCT TAACGTTTAA CGAAAGCCAT CTAACCTCTGT  
 AGCAAATGCT GGTTCGAAG CTGCTAAGGC TTTGACCGAT AACGCTCAAT 150  
 TCGTTTACGA CCAAACGTTT GACGATTCCG AACTGGCTA TTGCGAGTTA  
 CTTTGTTTAA TGGTGCTGCT CAAGCTGTTT ACAACAAGTT CCATACACC 200  
 GAAACCAATT ACCACGACGA GTTCGACAAA TGTGTTCAA GGGTATGTGG  
 ACTCAAACCC AAGGTAACAA CTTGCTGCA GATCAAAGAG GTAAGGATAA 250  
 TGAGTTTGGG TTCCATTGTT GAAGCGACGT CTAGTTTCTC CATTCCTATT  
 GTGTGCTAGA GATATTGGTT ACTACTTGAG AATTGTTACC TACTGTTTGG 300  
 CACACGATCT CTATAACCAA TGATGAACTC TTAACAATGG ATGACAAACC  
 TTGCAGGTGG TACTGGTCCA TTGGATGAAT ACTTGATTGC TGGTATTGAT 350  
 AACGTCCACC ATGACCAGGT AACCTACTTA TGAACCTAAG ACCATAACTA  
 GAAATTAACA GAACCTTCGA TTTGTCTCCA TCTTGGTACG TTGAAGCATT 400  
 CTTTAATTGT CTTGGAAGCT AAACAGAGGT AGAACCATGC AACTTCGTAA  
 GAAGTACATT AAGGCAAATC ATGGTTTATC TGGTGATGCT AGAGATGAAG 450  
 CTTTCATGTAA TTCCGTTTAG TACCAAATAG ACCACTACGA TCTCTACTTC  
 CAAACTCTTA CTTGATTAC GCTATTAACG CTTTGTCTAA CTCGAG 500  
 GTTTGAGAAT GAACCTAATG CGATAATTGC GAAACAGATT GAGCTC  
~~Seq. ID. 6~~

SEQ ID NO 7

TCATGAGTGA ACCAAACTTG AACCCAGCTT ACACCTTGGA TCAAGCTATT 50  
 AGTACTCACT TGGTTTGAAC TTGGGTCGAA TGTGGAACCT AGTTCGATAA  
 GCAAACCTTGC AACAAACCGA AGATGCTTCT GCTAGATACT ATGCTGCT TG 100  
 CGTTTGAACG TTGTTTGGCT TCTACGAAGA CGATCTATGA TACGACGAAC  
 GTGGATTGGT AGATTCAGAG CTGCTCAACC AGAAACCATT GCTGCTTTGT 150  
 CACCTAACCA TCTAAGTCTC GACGAGTTGG TCTTTGGTAA CGACGAAACA  
 TGGTTGCTTT GGAAGATGAA ACCGATAGAT CACCAGATGG TGGTTACCCA 200  
 ACCAACGAAA CCTTCTACTT TGGCTATCTA GTGGTCTACC ACCAATGGGT  
 TTGAGAAGAA ACGCTGCTAA GGCTTTGGGT AAATTGGGTG ATAGACAAGT 250  
 AACTCTTCTT TCGGACGATT CCGAAACCCA TTAAACCCAC TATCTGTTCA  
 TGTTCACGCT TTGATTAAGG CTTTGGGAATG TGAAGATTAC TACGTTAGAG 300  
 ACAAGGTCGA AACTAATTCC GAAACCTTAC ACTTCTAATG ATGCAATCTC  
 AATCTGCTGC TCAAGCATTG GAAGGTTTGG GTGATGCTAG AGCTATGCT 350  
 TTAGACGACG AGTTCGTAAC CTTCCAAACC CACTACGATC TCGATACCGA  
 CCATTGATGG CTAAGTTGAC CGGTGGTTTG GCTGCTGCTC AATTGGT TGA 400  
 GGTAACCTACC GATTCAACTG GCCACCAAAC CGACGACGAG TTAACCAACT  
 AGGTAAGCCA CATTTGGCTC AACCATACGA AGCTATCATT GAAGCAT TGG 450  
 TCCATTCCGT GTAAACCGAG TTGGTATGCT TCGATAGTAA CTTGTAACC  
 GTACTTTGCA AGCTGTTGAA TCTATTGGTT TGATTGAACC ATTCTTGAA 500  
 CATGAAACGT TCGACAACCT AGATAACCAA ACTAACTTGG TAAGAACCTT  
 CATTTCTCAC CAAAGTTTCA ATACGCTGCT GCTAGAGCTT TGTTCCAAAT 550  
 GTAAAGAGTG GTTTCCAAGT TATGCGACGA CGATCTCGAA ACAAGGTTAA  
 GACCGGTGAT AACAGATACG GTGATTTGTT GATTACCGCT TTGGGTGGTA 600  
 CTGGCCACTA TTGTCTATGC CACTAAACAA CTAATGGCGA AACCACCAT  
 CAGATTTGCA ATTGAGAAGA TCAGCTATGA TGGATTTGGG TGCTACTGGT 650  
 GTCTAAACGT TAACTCTTCT AGTCGATACT ACCTAAACCC ACGATGACCA  
 TACTTACCAG GTGCTCAAGC TATTGCTAAG GCTTTCGCTG AAAACT CTTT 700  
 ATGAATGGTC CACGAGTTTCG ATAACGATTG CGAAAGCGAC TTTTGA.GAAA  
 GAAGTTGATT GCTTTGAGAG ATTTGTGGGC TACCCATAGA CAAAGACAAG 750  
 CTTCAACTAA CGAAACTCTC TAAACACCCG ATGGGTATCT GTTTCTGTTC  
 CATCTTCTGA ATCTAAGGCT TTGTCTCCAG CTTCAAGACA AATTTTGGAA 800  
 GTAGAAGACT TAGATTCCGA AACAGAGGTC GAAGTTCTGT TAAAAACCTT  
 TTGATGGATT CTTTGTGAA CTCGAG 850  
 AACTACCTAA GAAACAACCT GAGCTC  
 -XhoI-

Seq ID No. 8

<sup>Start</sup>  
 CCATGGAAGG TAACTCTGTT GTTACCCAG AAATTGAAAG ATTGATTCAA 50  
 GGTACCTTCC ATTGAGACAA CAATGGGGTC TTTAACTTTC TAACTAAGTT  
 GCTGTTGAAA CCGCAGATTC TGCTGCTAAG TTAGTTGGTG CTGTTAGAGC 100  
 CGACAACCTT GCGCTCTAAG ACGACGATTC AATCAACCAC GACAATCTCG  
 TTTGGCTGCT ACCAGATCAC CATTGGCTGT TCCACAATTG ACCACCGTTT 150  
 AAACCGACGA TGGTCTAGTG GTAACCGACA AGGTGTTAAC TGGTGGCAAA  
 TGAGATACAA CAACCCAGGT GCTGCTGTTG CTGCAGTTGA TGGTTTGATT 200  
 ACTCTATGTT GTTGGGTCCA CGACGACAAC GACGTCAACT ACCAAACTAA  
 CAAATTGGTG ATGCTGCTAT GACCCATTG TTGGCAAACA TGGATGGTTA 250  
 GTTTAACCAC TACGACGATA CTGGGTAAAC AACCGTTTGT ACCTACCAAT  
 CAACTACGGT GCTAGAGCTT GGGCTACTAG AGCTTGTGCT GGTATTGGTG 300  
 GTTGATGCCA CGATCTCGAA CCCGATGATC TCGAACACGA CCATAACCAC  
 ATCCAAGAGC TTTGGCTTTG TTGCAAGAAG CTGCTTTGAC CGATTTCCGT 350  
 TAGGTTCTCG AAACCGAAAC AACGTTCTTC GACGAACTG GCTAAAGCGA  
 TTGTCTGTTA GAAGAGCTGC TGCTAAGGGT TTGGGTTTCT TGAGATGGCA 400  
 AACAGACAAT CTTCTCGACG ACGATTCCCA AACCCAAAGA ACTCTACCGT  
 ATCTTTGCCA CAAGAAGAAC AAGAAACCGT TCAAAAGGCT ATTTACGATA 450  
 TAGAAACGGT GTTCTTCTTG TTCTTTGGCA AGTTTCCGA TAAATGCTAT  
 CCTTGATTCA AGTTTGTGAA GATCCAGAAT GGGTTGTTAG ATACGGTGCT 500  
 GGAACATAAGT TCAAACACTT CTAGGTCTTA CCCAACAATC TATGCCACGA  
 ATTGCTGGTT TGGAAACTT GGCTAAGCAA GCTCAACATT ACAGACAACC 550  
 TAACGACCAA ACCTTTTGAA CCGATTTCGT CGAGTTGTAA TGTCTGTTGG  
 ATTGAAGGAT TTCTTGCAAT CTTTCGTTGA ACAAGAACCA GAAGCTATTG 600  
 TAACTTCCTA AAGAACGTTA GAAAGCAACT TGTTCCTGGT CTTCGATAAC  
 TTGGTGAAAG AATTTTGTGG ACCTTGGAAG ACATTGGTCC AATTA<sup>Stop</sup>ACTCG 650  
 AACCCTTTC TTTAAACACC TGGAACCTTT TGTAAACCAG TTAATTGAGC  
 AG 700  
 TC

Seq. ID. No 9

Expt 17  
 TCATGAGTGT AAAC TTGGCT TCACAATTGA GAGAAGGTAC TAAGAAGTCT 50  
 AGTACTCACA TTGGAACCGA AGTGTTAACT CTCTTCCATG ATTCTTCAGA  
 CATTCTATGG CTGAAAACGT TGGTTTCGTT AAGTGTTTCT TGAAGGGTGT 100  
 GTAAGATACC GACTTTTGCA ACCAAAGCAA TTCACAAAGA ACTTCCCACA  
 TGTGAAAAG AACTCTTACA GAAAGTTAGT TGGTAACTTG TACTTCGTTT 150  
 ACAACTTTTC TTGAGAATGT CTTTCAATCA ACCATTGAAC ATGAAGCAAA  
 ACTCTGCTAT GGAAGAAGAA ATGGCTAAGT TCAAGGATCA TCCAATTTTG 200  
 TGAGACGATA CCTTCTTCTT TACCGATTCA AGTTCTTAGT AGGTTAAAC  
 TCTCATATCT ACTTCCCAGA ATTGAACAGA AAGCAATCTT TGAACAAGA 250  
 AGAGTATAGA TGAAGGGTCT TAACTTGTCT TTCGTTAGAA ACCTTGTCT  
 TTTGCAATTC TACTACGGTT CAACTGGAG ACAAGAAGTT AAGATTTCTG 300  
 AAACGTTAAG ATGATGCCAA GTTTGACCTC TGTTCTTCAA TTCTAAAGAC  
 CTGCTGGTCA AGCATACGTT GATAGAGTTA GACAAGTTGC TGCTACCGCT 350  
 GACGACCACT TCGTATGCAA CTATCTCAAT CTGTTCAACG ACGATGGCGA  
 CCAGAATTGT TGGTTGCTCA TTCTTACACC AGATACTTGG GTGATTTGTC 400  
 GGTCTTAACA ACCAACGAGT AAGAATGTGG TCTATGAACC CACTAAACAG  
 TGGTGGTCAA ATTTTGAAGA AGATTGCTCA AAACGCTATG AACTTGCATG 450  
 ACCACCAGTT TAAACTTCT TCTAACGAGT TTTGCGATAC TTGAACGTAC  
 ATGGTGGTAC TGCTTTCTAC GAATTTGCAG ATATTGATGA TGAAGGCT 500  
 TACCACCATG ACGAAGATG CTTAAACGTC TATACTACT ACTTTCCGA  
 TTCAAGAACA CCTACAGACA AGCTATGAAC GATTTGCCAA TTGATCAAGC 550  
 AAGTTCTTGT GGATGTCTGT TCGATACTTG CTAAACGGTT AACTAGTTCG  
 TACCGCTGAA AGAATTGTTG ATGAAGCAAA CGATGCTTTC GCTATGAACA 600  
 ATGGCGACTT TCTTAACAAC TACTTCGTTT GCTACGAAAG CGATACTTGT  
 TGAAGATGTT CAACGAATTG GAAGGTAAC TGAATTAAGG TATTGGTATT 650  
 ACTTCTACAA GTTGCTTAAC CTTCCATTGA ACTAATCCG ATAACCATAA  
 ATGGTTTCA ACTCTTGAC CAGAAGAAGA TCACAAGGTT CTACCGAAGT 700  
 TACCAAAAGT TGAGAACTG GTCTTCTTCT AGTGTTCCAA GATGGCTTCA  
 TGGTTTGGCT ACCTCTGAAG GTAACGCGAG 750  
 ACCAAACCGA TGGAGACTTC CATTGAGCTC  
 ————— /

Seq. ID No 10

~MeI  
 CCATGGCTGT TACCGATTTG TCTTTGACCA ACTCTTCTTT GATGCCAACC 50  
 GGTACCGACA ATGGCTAAAC AGAAACTGGT TGAGAAGAAA CTACGGTTGG  
 TTGAACCCAA TGATTCAACA ATTGGCTTTG GCTATTGCTG CTTCTTGGCA 100  
 AACTTGGGTT ACTAAGTTGT TAACCGAAAC CGATAACGAC GAAGAACCGT  
 ATCTTTGCCA TTGAAGCCAT ACCAATTGCC AGAAGATTTG GGTTACGTTG 150  
 TAGAAACGGT AACTTCGGTA TGGTTAACGG TCTTCTAAAC CCAATGCAAC  
 AAGGCAGATT GGAAGGTGAA AAGTTGGTTA TTGAAAACAG ATGTTACCAA 200  
 TTCCGTCTAA CCTTCCACTT TTCAACCAAT AACTTTTGTC TACAATGGTT  
 ACCCCACAAT TCAGAAAGAT GCATTTGGAA TTGGCTAAAG TTGGTAAGGG 250  
 TGGGGTGTTA AGTCTTTCTA CGTAAACCTT AACCGATTTC AACCATTCCC  
 TTTGGATATT TTGCATTGTG TTATGTTCCC AGAACCATTG TACGGTTTGC 300  
 AAACCTATAA AACGTAACAC AATACAAGGG TCTTGGTAAC ATGCCAAACG  
 CATTGTTTCG TTGTGATATT GTTGCTGGTC CAGGTGGTGT TTCTGCTGCT 350  
 GTAACAAGCC AACACTATAA CAACGACCAG GTCCACCACA AAGACGACGA  
 ATTGCAGATT TGTCTCCAAC CCAATCAGAT AGACAATTGC CAGCTGCTTA 400  
 TAACGTCTAA ACAGAGGTTG GGTTAGTCTA TCTGTTAACG GTCGACGAAT  
 CCAAAAGTCT TTGGCTGAAT TGGGTCAACC AGAATTTGAA CAACAAAGAG 450  
 GGTTTTCAGA AACCGACTTA ACCCAGTTGG TCTTAACTT GTTGTTCCTC  
 AATTGCCACC TTGGGGTGAA ATTTTCTCTG AATACTGTTT GTTCATTAGA 500  
 TTAACGGTGG AACCCCACTT TAAAAGAGAC TTATGACAAA CAAGTAATCT  
 CCATCAAACG TTACCGAAGA AGAAAGATTC GTTCAAAGAG TTGTTGATTT 550  
 GGTAGTTTGC AATGGCTTCT TCTTTCTAAG CAAGTTTCTC AACAACTAAA  
 CTTGCAAATT CATTGTCATC AATCTATTGT TGCTGAACCA TTGTCTGAAG 600  
 GAACGTTTTA GTAACAGTAG TTAGATAACA ACGACTTGGT AACAGACTTC  
 CTCAACTTT GGAACATAGA CAAGGTCAAA TTCATTACTG TCAACAACAA 650  
 GAGTTTGAAA CCTTGTATCT GTTCCAGTTT AAGTAATGAC AGTTGTTGTT  
 CAAAAGAACG ATAAGACCAG AAGAGTTTTG GAAAAGGCTT TCGGTGAAGC 700  
 GTTTTCTTGC TATTCTGGTC TTCTCAAAAC CTTTTCCGAA AGCCACTTCG  
 ATGGGCTGAA AGATACATGT CTCAAGTTTT GTTCGATGTT ATTCAAACT 750  
 TACCCGACTT TCTATGTACA GAGTTCAAA CAAGCTACAA TAAGTTTGA  
 ~MeI  
 CGAG 800  
 GCTC

SEQ. ID. No 11

ATGTTGACGTATTCACCTCGGGTTGTTTCCCAAGCTGATGCTCGCGGCGAGTACCT  
CTCTGGTTCTCAGTTAGATGCTTTGAGCGCTACCGTTGCTGAAGGCAACAAACGG  
ATTGATTCTGTAAACCGCATCACCGGTAATGCTTCCGCTATCGTTTCCAACGCTGC  
TCGTGCTTTGTTTCGTTGAACAGCCCCAATTAATCCAACCCGGTGGAACGCCTAC  
ACCAGCCGTCGTATGGCTGCTTGTTTGCGTGACATGGAAATCATCCTCCGCTATGT  
TACCTACGCAACCTTCACCGGCGACGCTTCCGTTCTAGAAGATCGTTGCTTGAAC  
GGTCTCCGTGAAACCTACGTTGCCCTGGGTGTTCCCGGTGCTTCCGTAGCTGCTGG  
CGTTCAAAAAATGAAAGAAGCTGCCCTGGACATCGTTAACGATCCCAATGGCATC  
ACCCGTGGTGATTGCAGTGCTATCGTTGCTGAAATCGCTGGTTACTTCGACCGCG  
CCGCTGCTGCCGTAGCCTAG

SEQ. ID No 12

MFDVFTRVVSQADARGEYLSGSQLDALSATVAEGNKRIDSVNRITGNASAIVSNAAR  
ALFVEQPQLIQPGGNAYTSRRMAACLRDMEILLRYVTYATFTGDASVLEDRCNLGLR  
ETTYVALGVPGASVAAGVQKMKEAALDIVNDPNGITRGDCSAIVAEIAGYFDRAAAA  
VA

TCGCGCGTTTCGGTGATGACGGTGAAAAACCTCTGACACATGCAGCTCCCGGAGAC  
GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC  
GTCAGCGGGTGTGGCGGGTGTGCGGGCTGGCTTAACCTATGCGGCATCAGAGCAG  
ATTGTAAGTGCAGAGTGCACACGCTTTTCAATTCAATTCATCATTTTTTTTTTATTCT  
TTTTTTTGATTTTCGGTTTCTTTGAAATTTTTTTTGATTTCGGTAATCTCCGAACAGAAG  
GAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACGCATATGTAGT  
GTTGAAGAAACATGAAATTGCCAGTATTCTTAACCCAACTGCACAGAACAAAA  
ACCTGCAGGAAACGAAGATAAATCATGTGCGAAAGCTACATATAAGGAACGTGCT  
GCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGC  
AAACAAACTTGTGTGCTTCATTGGATGTTTCGTACCACCAAGGAATTACTGGAGTT  
AGTTGAAGCATTAGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTG  
ACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGT  
ACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAA  
ATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAAT  
GCACACGGTGTGGTGGGCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCAGAA  
GAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAAATTGTCATGCAAG  
GGCTCCCTATCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCG  
ACAAAGATTTTGTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGA  
AGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGAC  
GCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACA  
TTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTACAGG  
GTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGAGCAGC  
AAAACCTAAAAAACTGTATTATAAGTAAATGCATGTATACTAACTCACAAATTAG  
AGCTTCAATTTAATTATATCAGTTATTACCCTGCGGTGTGAAATACCGCACAGAT  
GCGTAAGGAGAAAAATACCGCATCAGGAAATTGTAAACGTTAATATTTTGTAAAA  
TTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGG  
CAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTGTTTCCA  
GTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA  
AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTT  
TTTTGGGGTTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCC  
GATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG  
AAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGC  
GTAACCACACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCGCGCCATT  
CGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCT  
ATTACGCCAGCTGGCGAAGGGGGGATGTGCTGCAAGGCGATTAAAGTTGGGTAAC  
GCCAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATAC  
GACTCACTATAGGGCGAATTGGAGCTCATATCCTTTTGTGTTTCCGGGTGTACAA  
TATGGACTTCCTCTTTTCTGGCAACCAAACCCATACATCGGGATTCTATAATACC  
TTCGTTGGTCTCCCTAACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGA  
TACCAGACAAGACATAATGGGCTAAACAAGACTACACCAATTACACTGCCTCATT  
GATGGTGGTACATAACGAACATAACTGTAGCCCTAGACTTGATAGCCATCATCA  
TATCGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAGTAATAATAGGCG  
CATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATA  
TCCGCAATGACAAAAAAATGATGGAAGACACTAAAGGAAAAAAATTAAACGACAAA  
GACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGGGGTATCTCGAAGCAC  
ACGAAACTTTTTCTTCTCCTTCATTACGCACACTACTCTCTAATGAGCAACGGTAT  
ACGGCCTTCTTCCAGTTACTTGAATTTGAAATAAAAAAAGTTTGCTGTCTTGCT  
ATCAAGTATAAATAGACCTGCAATTATTAATCTTTTGTTCCTCGTCATTGTTCTC  
GTTCCCTTTCTCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATA



CAATCAACTATCTCATATACACCATGGAAGGTAACCTCTGTTGTTACCCCAGAAAT  
TGAAAGATTGATTCAAGCTGTTGAAACCGCAGATTCTGCTGCTAAGTTAGTTGGT  
GCTGTTAGAGCTTTGGCTGCTACCAGATCACCATTGGCTGTTCCACAATTGACCAC  
CGTTTTGAGATACAACAACCCAGGTGCTGCTGTTGCTGCAGTTGATGGTTTGATT  
AAATTGGTGATGCTGCTATGACCCATTTGTTGGCAAACATGGATGGTTACAACATA  
CGGTGCTAGAGCTTGGGCTACTAGAGCTTGTGCTGGTATTGGTGATCCAAGAGCT  
TTGGCTTTGTTGCAAGAAGCTGCTTTGACCGATTTCGCTTTGTCTGTTAGAAGAGC  
TGCTGCTAAGGGTTTGGGTTTCTTGAGATGGCAATCTTTGCCACAAGAAGAACAA  
GAAACCGTTCAAAAGGCTATTTACGATACCTTGATTCAAGTTTGTGAAGATCCAG  
AATGGGTTGTTAGATACGGTGCTATTGCTGGTTTGGAAAACCTGGCTAAGCAAGC  
TCAACATTACAGACAACCATTTGAAGGATTTCTTGCAATCTTTCGTTGAACAAGAA  
CCAGAAGCTATTGTTGGTGAAAGAATTTGTGGACCTTGGAAAACATTGGTCCAA  
TTAACTCGAGATAAGGTATATAACTCTGTAGAAATAAAGAGTATCATCTTTCAA  
CCGCGGATATCCTTTTGTGTTTCCGGGTGTACAATATGGACTTCCTCTTTTCTGG  
CAACCAAAACCCATACATCGGGATTCTATAATACTTCGTTGGTCTCCCTAACATG  
TAGGTGGCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACATAATGG  
GCTAAACAAGACTACCAATTACACTGCCTCATTGATGGTGGTACATAACGAAC  
TAATACTGTAGCCCTAGACTTGATAGCCATCATCATATCGAAGTTTCACTACCCTT  
TTTCCATTGCCATCTATTGAAGTAATAATAGGCGCATGCAACTTCTTTTCTTTTT  
TTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAATG  
ATGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCG  
TTGTTCCAAGAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCCTTCCTC  
ATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCAGTTACT  
TGAATTTGAAATAAAAAAAGTTTGTCTGCTTGTCTATCAAGTATAAATAGACCTG  
CAATTATTAACTTTTGTCTTCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTCT  
TTTTCTGCACAATATTTCAAGCTATACCAAGCATAACAATCAACTATCTCATATACA  
CCATGGCTGTTACCGATTGTCTTTGACCAACTCTTCTTGATGCCAACCTTGAAC  
CCAATGATTCAACAATTGGCTTTGGCTATTGCTGCTTCTTGGCAATCTTGGCATT  
GAAGCCATACCAATTGCCAGAAGATTTGGGTTACGTTGAAGGCAGATTGGAAGG  
TGAAAAGTTGGTTATTGAAAACAGATGTTACCAAACCCCAATTCAGAAAAGATG  
CATTTGGAATTGGCTAAAGTTGGTAAGGGTTTGGATATTTGCATTGTGTTATGTT  
CCCAGAACCATTTGTACGGTTTGCCATTGTTTCGGTTGTGATATTGTTGCTGGTCCAG  
GTGGTGTCTCTGCTGCTATTGCAGATTTGTCTCCAACCAATCAGATAGACAATTG  
CCAGTGCTTACCAAAAGTCTTTGGCTGAATTGGGTCAACCAGAATTTGAACAAC  
AAAGAGAATTGCCACCTTGGGGTGAAATTTCTCTGAATACTGTTTGTTCATTAGA  
CCATCAAACGTTACCGAAGAAGAAAGATTTCGTTCAAAGAGTTGTTGATTTCTTGC  
AAATTCATTGTCATCAATCTATTGTTGCTGAACCATTGTCTGAAGCTCAAACCTTG  
GAACATAGACAAGGTCAAATTCATTACTGTCAACAACAACAAAAGAACGATAAG  
ACCAGAAAGAGTTTGGAAAAGGCTTTCGGTGAAGCATGGGCTGAAAGATACATG  
TCTCAAGTTTGTTCGATGTTATTCAAAACTCGAGATAAGGTATATAACTCTGTAG  
AAATAAAAGATATCATCTTTCAAACCGCGGATTGTGCGGATCAAATCGATATGTC  
TTATGCGGCCCGCATATCCTTTTGTGTTTCCGGGTGTACAATATGGACTTCCTCTTT  
TCTGGCAACCAAACCCATACATCGGGATTCTATAATACTTCGTTGGTCTCCCTA  
ACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACAT  
AATGGGCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTACATAA  
CGAACTAATACTGTAGCCCTAGACTTGATAGCCATCATCATATCGAAGTTTCACT  
ACCCTTTTCCATTGTCATCTATTGAAGTAATAATAGGCGCATGCAACTTCTTTT  
CTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAA  
AAAATGATGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACA  
GATGTCGTTGTTCCAGAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTC  
CTTCCTTCATTACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTC

CAGTTACTTGAATTTGAAATAAAAAAAGTTTGCTGTCTTGCTATCAAGTATAAA  
TAGACCTGCAATTATTAATCTTTTGTTCCTCGTCATTGTTCTCGTTCCTTTCTTC  
CTTGTTTCTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTATC  
TCATATACACCATGAGTGAACCAAACCTTGAACCCAGCTTACACCTTGGATCAAGC  
TATTGCAAACCTTGCAACAAACCGAAGATGCTTCTGCTAGATACTATGCTGCTTGG  
TGGATTGGTAGATTCAAGAGCTGCTCAACCAGAAACCATTGCTGCTTTGTTGGTTG  
CTTTGGAAGATGAAACCGATAGATCACCAGATGGTGGTTACCCATTGAGAAGAA  
ACGCTGCTAAGGCTTTGGGTAAATTGGGTGATAGACAAGTTGTTCCAGCTTTGAT  
TAAGGCTTTGGAATGTGAAGATTACTACGTTAGAGAATCTGCTGCTCAAGCATTG  
GAAGGTTTGGGTGATGCTAGAGCTATGGCTCCATTGATGGCTAAGTTGACCGGTG  
GTTTGGCTGCTGCTCAATTGGTTGAAGGTAAGCCACATTTGGCTCAACCATACGA  
AGCTATCATTGAAGCATTGGGTACTTTGCAAGCTGTTGAATCTATTGGTTTGATTG  
AACCATTCTTGGAACATTTCTCACCAAAGGTTCAATACGCTGCTGCTAGAGCTTTG  
TTCCAATTGACCGGTGATAACAGATACGGTGATTTGTTGATTACCGCTTTGGGTG  
GTACAGATTTGCAATTGAGAAGATCAGCTATGATGGATTTGGGTGCTACTGGTTA  
CTTACCAGGTGCTCAAGCTATTGCTAAGGCTTTTCGCTGAAAACCTTTTGAAGTTGA  
TTGCTTTGAGAGATTGTTGGGCTACCCATAGACAAAGACAAGCATCTTCTGAATC  
TAAGGCTTTGTCTCCAGCTTCAAGACAAATTTTGGAATTGATGGATTCTTTGTTGA  
ACTCGAGATAAGGTATATACTCTGTAGAAATAAAGAGTATCATCTTTCAAACCG  
CGGATTGTCGCGATCAAATCGATATGTCTTATGCGGCCGCTTACGACCGTTAACTT  
GTTCTAGAATATCCTTTTGTGTTTCCGGGTGTACAATATGGACTTCCTCTTTCTG  
GCAACCAAACCCATACATCGGGATTCTATAATACCTTCGTTGGTCTCCCTAACAT  
GTAGGTGGCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACATAATG  
GGCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTACATAACGAA  
CTAATACTGTAGCCCTAGACTTGATAGCCATCATCATATCGAAGTTTCACTACCCT  
TTTTCCATTTGCCATCTATTGAAGTAATAATAGGCGCATGCAACTTCTTTTCTTTT  
TTTTCTTTTCTCTCTCCCCCGTTGTGTCTCACCATATCCGCAATGACAAAAAAT  
GATGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTC  
GTTGTTCCAGAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCTTCCTT  
CATTACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCCAGTTAC  
TTGAATTTGAATAAAAAAAGTTTGCTGTCTTGCTATCAAGTATAAATAGACCT  
GCAATTATTAATCTTTTGTTCCTCGTCATTGTTCTCGTTCCTTTCTTCTTGTTC  
TTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTATCTCATAC  
ACCATGCTGATAAAGTTGGCTTCACAATTGAGAGAAGGTACTAAGAAGTCTCATT  
CTATGGCTGAAAACGTTGGTTTCGTTAAGTGTTTCTTGAAGGGTGTGTTGAAAA  
GAACTCTTACAGAAAGTTAGTTGGTAACCTTGTACTTCGTTTACTCTGCTATGGAAG  
AAGAAATGGCTAAGTTCAAGGATCATCCAATTTTGTCTCATATCTACTTCCCAGA  
ATTGAACAGAAAGCAATCTTTGGAACAAGATTTGCAATTCTACTACGGTTCAAAC  
TGGAGACAAAGAGTTAAGATTTCTGCTGCTGGTCAAGCATACGTTGATAGAGTTA  
GACAAGTTGCTGCTACCGCTCCAGAATTGTTGGTTGCTCATTCTTACACCAGATAC  
TTGGGTGATTTGTCTGGTGGTCAAATTTGAAGAAGATTGCTCAAAACGCTATGA  
ACTTGCATGATGGTGGTACTGCTTTCTACGAATTTGCAGATATTGATGATGAAAA  
GGCTTTCAAGAACACCTACAGACAAGCTATGAACGATTTGCCAATTGATCAAGCT  
ACCGCTGAAAGAATTGTTGATGAAGCAAACGATGCTTTCGCTATGAACATGAAGA  
TGTTCAACGAATTGGAAGGTAACCTTGATTAAGGCTATTGGTATTATGGTTTCAAC  
TCTTTGACCAGAAGAAGATCACAAGGTTCTACCGAAGTTGGTTTGGCTACCTCTG  
AAGGTAACCTCGAGATAAGGTATATACTCTGTAGAAATAAAGAGTATCATCTTTC  
AAACCGCGGATTGTCGCGATCAAATCGATATGTCTTATGCGGCCGCTTACGACCG  
TTAACTTGTTCTAGATTGGGCTAGCGTTGAGATCTATATCCTTTTGTGTTTCCGG  
GTGTACAATATGGACTTCCTCTTTTCTGGCAACCAAACCCATACATCGGGATTCT  
ATAATACCTTCGTTGGTCTCCCTAACATGTAGGTGGCGGAGGGGAGATATACAAT

AGAACAGATACCAGACAAGACATAATGGGCTAAACAAGACTACACCAATTACAC  
TGCCTCATTGATGGTGGTACATAACGAACTAATACTGTAGCCCTAGACTTGATAG  
CCATCATCATATCGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAGTAA  
TAATAGGCGCATGCAACTTCTTTCTTTTTTTTTCTTTCTCTCTCCCCCGTTGTTGT  
CTCACCATATCCGCAATGACAAAAAATGATGGAAGACACTAAAGGAAAAAATT  
AACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGGGGTAT  
CTCGAAGCACACGAAACTTTTTCTTCCCTTCATTACGACACTACTCTCTAATGA  
GCAACGGTATACGGCCTTCCTTCCAGTTACTTGAATTTGAAATAAAAAAAGTTT  
GCTGTCTTGCTATCAAGTATAAATAGACCTGCAATTATTAATCTTTTGTTCCTCG  
TCATTGTTCTCGTTCCTTTCTTCCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTAT  
ACCAAGCATACAATCAACTATCTCATATACACCATGAAGACCCCATTGACCGAAG  
CTGTTTCTACCGCAGATTCTCAAGGTAGATTCTTGTCATCTACCGAATTGCAAATT  
GCTTTCGGTAGATTGAGACAAAGCAAATGCTGGTTTGCAAGCTGCTAAGGCTTTGA  
CCGATAACGCTCAATCTTTGGTTAATGGTGCTGCTCAAGCTGTTACAACAAGTTC  
CCATACACCACTCAAACCCAAGGTAACAACCTTCGCTGCAGATCAAAGAGGTAAG  
GATAAGTGTGCTAGAGATATTGGTTACTACTTGAGAATTGTTACCTACTGTTTGGT  
TGCAGGTGGTACTGGTCCATTGGATGAATACTTGATTGCTGGTATTGATGAAATT  
AACAGAACCTTCGATTTGTCTCCATCTTGGTACGTTGAAGCATTGAAGTACATTA  
AGGCAAATCATGGTTTATCTGGTGTGCTAGAGATGAAGCAAACCTTACTTGGA  
TTACGCTATTAACGCTTTGTCTAACTCGAGATAAGGTATATAACTCTGTAGAAAT  
AAAGAGTATCATCTTTCAAACCGCGGATTGTGCGGATCAAATCGATATGTCTTAT  
GCGGCCGCTTACGACCGTTAACTTGTTCTAGATTGGGCTAGCGTTGAGATCTTTAG  
AAACGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTCCCTTTAGTGAGG  
GTTAATTCCGAGCTTGCGGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTT  
ATCCGCTCACAATTCACACAACATAGGAGCCGGAAGCATAAAGTGTAAGCCT  
GGGGTGCCTAATGAGTGAGGTAACCTCACATTAATTGCGTTGCGCTCACTGCCCGC  
TTCCAGTCCGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCG  
GGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT  
GCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATA  
CGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGC  
CAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG  
CTCGGCCCCCTGACGAGCATCAGAAAAATCGACGCTCAAGTCAGAGGTGGCGA  
AACCCGACAGGACTATAAAGATACAGGCGTTCCCCCTGGAAGCTCCCTCGTGC  
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG  
GGAAGCGTGCGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGG  
TCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTG  
CGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCG  
CCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGT  
GCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTAT  
TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC  
TTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAG  
CAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGG  
GGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT  
ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA  
ATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG  
AGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTGCCC  
GTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAA  
TGATAACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCC  
AGCCGGAAGGGCCGAGCGCAGAAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAG  
TCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC  
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATG

GCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTT  
GTGAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAAGTAAGTTG  
GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT  
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGA  
GAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATA  
CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGG  
GCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACT  
CGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC  
AAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT  
GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTATT  
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAACAAATAGGGGT  
TCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATC  
ATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCCTTCGTC